DOCUMENT RESUME

ED 371 423 CS 508 602

AUTHOR	Fowler, Carol A., Ed.
TITLE	Status Report on Speech Research, April-June 1993. SR-114.
INSTITUTION	Haskins Labs., New Haven, Conn.
SPONS AGENCY	National Inst. of Child Health and Human Development (NIH), Bethesda, MD.
PUB DATE	93
CONTRACT	DBS-9112198; HD-01994; HD-21888
NOTE	192p.; For the January-March 1993 report, see ED 366 020.
PUB TYPE	Collected Works - General (020) -- Reports - Research/Technical (143)
EDRS PRICE	MF01/PC08 Plus Postage.
DESCRIPTORS	Adults; Chinese; Communication Research; Elementary Secondary Education; Higher Education; Infants; *Language Acquisition; Language Research; Listening Skills; Literacy; Music; *Speech Communication; Thai; Word Recognition
IDENTIFIERS	Speech Research

ABSTRACT

One of a series of quarterly reports, this publication contains 11 articles which report the status and progress of studies on the nature of speech, instruments for its investigation, and practical applications. Articles in the publication are: "Evidence of Flexible Coding in Visual Word Recognition" (Kenneth R. Pugh and others); "Horizontal and Vertical Views of Chinese Psycholinguistics" (Ignatius G. Mattingly); "Learning to Perceive the Sound Pattern of English" (Catherine T. Best); "The Perceptual Infrastructure of Early Phonological Development" (Alice Faber and Catherine Best); "The Stability of Distinctive Vowel Length in Thai" (Arthur S. Abramson); "Calibration, Validation, and Hardware-Software Modifications to the Carstens EMMA System" (Peter J. Alfonso and others); "Morphological Analysis and the Acquisition of Morphology and Syntax in Specifically-Language-Impaired Children" (Karen M. Smith-Lock); "Hemispheric Asymmetries in Adults' Perception of Infant Emotional Expressions" (Catherine T. Best and others); "On Determining the Basic Tempo of an Expressive Music Performance" (Bruno H. Repp); "Musical Motion: Some Historical and Contemporary Perspectives" (Bruno H. Repp); and "A Review of P. Downing, S. D. Lima, & M. Noonan (Eds.), 'The Linguistics of Literacy'" (Ignatius G. Mattingly). (RS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 371 423

CS

Haskins Laboratories Status Report on

Speech Research

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CS508602

SR-114
APRIL-JUNE 1993

BEST COPY AVAILABLE

Haskins
Laboratories
Status Report on

Speech Research

SR-114
APRIL-JUNE 1993

NEW HAVEN, CONNECTICUT

Distribution Statement

Editor
Carol A. Fowler

Production
Yvonne Manning
Fawn Zefang Wang

This publication reports the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications.

Distribution of this document is unlimited. The information in this document is available to the general public. Haskins Laboratories distributes it primarily for library use. Copies are available from the National Technical Information Service or the ERIC Document Reproduction Service. See the Appendix for order numbers of previous Status Reports.

Correspondence concerning this report should be addressed to the Editor at the address below:

Haskins Laboratories
270 Crown Street
New Haven, Connecticut
06511-6695

Phone: (203) 865-6163 *FAX:* (203) 865-8963 *Bitnet:* HASKINS@YALEHASK
Internet: HASKINS%YALEHASK@VENUS.YCC.YALE.EDU

♻ This Report was reproduced on recycled paper ♻

Acknowledgment

The research reported here was made possible in part by support from the following sources:

National Institute of Child Health and Human Development

Grant HD-01994
Grant HD-21888

National Institute of Health

Biomedical Research Support Grant RR-05596

National Science Foundation

Grant DBS-9112198

National Institute on Deafness and Other Communication Disorders

Grant DC 00121
Grant DC 00183
Grant DC 00403
Grant DC 00016
Grant DC 00594
Grant DC 00865
Grant DC 01147
Grant DC 00044
Grant DC 00825
Grant DC 01247

Investigators

Arthur Abramson*
Peter J. Alfonso*
Eric Bateson*
Fredericka Bell-Berti*
Catherine T. Best*
Susan Brady*
Catherine P. Browman
Claudia Carello*
Franklin S. Cooper*
Stephen Crain*
Lois G. Dreyer*
Alice Faber
Laurie B. Feldman*
Janet Fodor*
Anne Fowler*
Carol A. Fowler*
Louis Goldstein*
Carol Gracco
Vincent Gracco
Katherine S. Harris*
John Hogden
Leonard Katz*
Rena Arens Krakow*
Andrea G. Levitt*
Alvin M. Liberman*
Diane Lillo-Martin*
Leigh Lisker*
Anders Löfqvist
Ignatius G. Mattingly*
Nancy S. McGarr*
Richard S. McGowan
Patrick W. Nye
Kiyoshi Oshima†
Kenneth Pugh*
Lawrence J. Raphael*
Bruno H. Repp
Hyla Rubin*
Philip E. Rubin
Elliot Saltzman
Donald Shankweiler*
Jeffrey Shaw
Michael Studdert-Kennedy*
Michael T. Turvey*
Douglas Whalen

Technical Staff

Michael D'Angelo
Vincent Gulisano
Donald Hailey
Yvonne Manning
William P. Scully
Fawn Zefang Wang
Edward R. Wiley

Administrative Staff

Philip Chagnon
Alice Dadourian
Betty J. DeLise
Lisa Fresa
Joan Martinez

Students*

Melanie Campbell
Sandra Chiang
Margaret Hall Dunn
Terri Erwin
Joseph Kalinowski
Laura Koenig
Betty Kollia
Simon Levy
Salvatore Miranda
Maria Mody
Weijia Ni
Mira Peter
Christine Romano
Joaquin Romero
Dorothy Ross
Arlyne Russo
Michelle Sancier
Sonya Sheffert
Caroline Smith
Brenda Stone
Mark Tiede
Qi Wang
Yi Xu
Elizabeth Zsiga

---

*Part-time
†Visiting from University of Tokyo, Japan

Contents

Evidence of Flexible Coding in Visual Word Recognition
Kenneth R. Pugh, Karl Rexer, and Leonard Katz 1

Horizontal and Vertical Views of Chinese Psycholinguistics
Ignatius G. Mattingly 27

Learning to Perceive the Sound Pattern of English
Catherine T. Best 31

The Perceptual Infrastructure of Early Phonological Development
Alice Faber and Catherine Best 81

The Stability of Distinctive Vowel Length in Thai
Arthur S. Abramson 91

Calibration, Validation, and Hardware-Software Modifications to the Carstens EMMA System
Peter J. Alfonso, J. Robert Neely, P. H. H. M. van Lieshout, W. Hulstijn, and H .F. M. Peters 101

Morphological Analysis and the Acquisition of Morphology and Syntax in Specifically-Language-Impaired Children
Karen M. Smith-Lock 113

Hemispheric Asymmetries in Adults' Perception of Infant Emotional Expressions
Catherine T. Best, Jane S. Womer, and Heidi Fréya Queen 139

On Determining the Basic Tempo of an Expressive Music Performance
Bruno H. Repp 159

Musical Motion: Some Historical and Contemporary Perspectives
Bruno H. Repp 167

A Review of P. Downing, S. D. Lima, & M. Noonan (Eds.), *The Linguistics of Literacy*
Ignatius G. Mattingly 179

Appendix 185

Haskins
Laboratories
Status Report on

Speech Research

Haskins Laboratories Status Report on Speech Research
1993, SR-114, 1-26

Evidence of Flexible Coding in Visual Word Recognition*

Kenneth R. Pugh,† Karl Rexer,‡ and Leonard Katz‡

Three visual word recognition experiments examined subjects' differential dependence on the phonological versus orthographic information in accessing the lexicon. The critical manipulation was the presence or absence of pseudohomophones (e.g., BOTE) in the nonword context of a lexical decision task. Subjects received either a list with no pseudohomophones (NPsH group) or between 17% and 30% pseudohomophones among the nonwords (PsH group). Performance on common set of words was contrasted. In the first experiment, subjects in the PsH group were faster and no less accurate than subjects in the NPsH group on word trials. Further, performance in the NPsH group was adversely affected by phonological inconsistency in the target's orthographic neighborhood while performance in the PsH group was not, suggesting that performance in the latter group depended less on the phonological route. In a second experiment, speed and accuracy advantages were once again obtained in the PsH condition on both the memory probe and lexical decision components of a dual-task paradigm. Neighborhood phonological inconsistency, once again, influenced the performance of only the NPsH group. In the final experiment a double lexical decision paradigm was employed. Relations among members of the word pairs were varied and included orthographically and phonologically similar pairs, orthographically similar but phonologically dissimilar pairs, and semantically related pairs. Subjects in the NPsH condition were adversely affected by phonological dissimilarity whereas PsH subjects were actually facilitated on these pairs. These results are consistent with the idea that the role of phonological processing varies as a function of experimental context.

The question of whether access to the mental lexicon during reading is mediated by phonological codes, visual codes, or both, has been debated and researched extensively in the field of cognitive psychology (Besner & Smith, 1992; Carello, Turvey & Lukatela, 1992; Carr & Pollatsek, 1985; Coltheart, Davelaar, Jonasson, & Besner, 1977; Humphreys & Evett, 1985; Rayner & Pollatsek, 1989; Seidenberg & McClelland, 1989; Van Orden, Pennington, & Stone, 1990). While it is widely accepted that both types of codes can be computed by a reader, a major issues concern whether one or the other type of coding will dominate the process, depending on factors such as word frequency, spelling regularity, reading experience, and type of orthography (Katz & Frost, 1992; Seidenberg, 1992; Van Orden et al., 1990; Waters & Seidenberg, 1985). An issue related to all of these factors is whether the relative contribution of each of these codes can be modulated by task demands. In several studies subjects have demonstrated an apparent flexibility in their degree of dependence on phonological or visual codes as a function of changes in experimental context; depending on task demands, phonological coding could be made either advantageous or disadvantageous, and subjects appeared to vary their behavior accordingly (Andrews, 1982; Hanson & Fowler, 1987; Hawkins, Reicher, Rogers, & Peterson, 1976; McQuade, 1981, 1983; Monsell, Patterson, Graham, Hughes, & Milroy, 1992; Paap & Noel, 1991; Shulman, Hornak, & Sanders, 1978). If

---

This research was supported by National Institute of Child Health and Human Development grant HD - 01994 to Haskins Laboratories. Mira Peter assisted us in running subjects, along with students in the first author's Seminar in Cognition.

adult readers do, in fact, possess the ability to change their dependence readily between phonological and visual codes, it is a matter of interest because it suggests that this flexibility is useful for their normal everyday reading; otherwise, why would a readiness for strategic variation exist at all? It is to this question of coding flexibility that the current research is addressed.

The Dual Route Debate

Research on letter string pronunciation has been strongly influenced by the idea that more than one way of generating a phonological output must exist in order to account for people's ability to pronounce both words which the reader has never seen before (including pseudowords, such as BINT) and words with exceptional or unconventional spelling-to-sound relations (e.g., AISLE and PINT). The speed with which subjects can name novel words or pseudowords suggests a process of early and efficient conversion from graphemic to phonologic codes; this compiled or *assembled* phonology may play a role in skilled reading of familiar words as well. The ability to correctly pronounce words that violate typical grapheme to phoneme conversion rules (e.g., PINT) suggests a lexical constraint on phonological output, and has been interpreted as evidence that phonological information can be recovered from lexicon: this information is called *addressed* phonology.

By far, the most popular way of coping with these considerations has been the so called dual-route theory of reading (Coltheart et al., 1977). Two routes to pronunciation are posited; a phonologic route and a direct access route (note that we use the term coding to describe the cognitive operations within these routes or pathways). The phonologic route is said to consist of two stages. One converts orthographic representations such as letters and letter clusters into appropriate phonological representations such as phonemes (assembled phonology). In a second stage, these phonological representations are matched to their appropriate lexical entries or, in the case of naming, to an appropriate articulation. The direct access route, on the other hand, is thought to involve direct mapping from orthographic representations to lexical entries. Although specific versions of dual route theories may differ on some point or another, the following assumptions are usually made explicitly or implicitly. First, that the two routes to lexicon, direct and phonologic, operate independently of one another. Second, given that the phonologic process logically requires an extra step, it will, on average, take longer to finish than direct access (Coltheart, 1978; Waters & Seidenberg, 1985; but see Stone & Van Orden, in press). Third, it is also assumed that as reading ability develops (or in the case of specific words, as familiarity increases) subjects will tend to bypass the phonological route and rely on the direct route for lexical access. (See Van Orden et al., 1990, for a detailed criticism of each of these assumptions.)

While the dual-route concept has continued to frame much of the experimental work in the word recognition field, all of its major tenets have been challenged in recent years (see Humphreys & Evett, 1985 and Van Orden et al. 1990 for reviews). The existence of context independent grapheme-to-phoneme conversion (GPC) rules has been challenged (Glushko, 1979; Humphreys & Evett, 1985). Empirical challenges come from what have been called consistency effects, wherein two words, both of which follow GPC rules, behave differently in a naming task if one of them has a neighbor that shares the target's orthographic rime but whose pronunciation of this rime is different than the target's (e.g., words like PINT and LINT). This effect suggests a lexical constraint on phonological mapping; pronunciation is strongly influenced by lexically stored information. However, a GPC process that is sensitive to frequency of occurrence and number of alternatives could be seen as consistent with these effects. In fact, Rosson (1985) has obtained evidence that words and nonwords with stronger rules (as indexed by the frequency of occurrence of their GPC mappings relative to others) are named more quickly than words with weaker rules, even when consistency effects are controlled for. This finding, while consonant with the GPC view, suggests that the process is sensitive to what Van Orden and his colleagues have termed "statistical regularity" between print and pronunciation (Van Orden et al., 1990).

Dual route accounts usually assume that phonological information builds up too slowly to be relevant to the processing of all but very low frequency words. That assumption has been challenged by data suggesting that phonological masking benefits target processing relative to orthographic masking even when the target is masked very shortly after presentation (Perfetti, Bell & Delaney, 1988; Perfetti & Bell, 1991). Further, the idea that phonological processing can be bypassed is challenged by Van Orden's categorization experiments (Van Orden, 1987; Van Orden, Johnston, & Hale, 1988). In these studies false positive error rates to homophone and pseudohomophone foils are much higher than to orthographically matched controls, suggesting an early influence of phonology on access to meaning. However, Jared and Seidenberg (1991) were able to replicate Van Orden's results only for low frequency words, and this finding is broadly consistent with dual-route accounts. In these models the visual route will tend to be slower for low frequency words and therefore greater phonological influences are possible on these items.

Lukatela and Turvey (1993) contrasted the naming latencies of high and low frequency words with their pseudohomophone counterparts (e.g., DOOR vs. FOAL and DORE vs. FOLE) under different levels of attentional load. They found that words and their pseudohomophones were similarly influenced by load, suggesting that they were processed in the same way. According to dual-route theory, words, especially high frequency ones, should be processed by the automatic direct route; therefore, dual-route theory had predicted interactions between lexical status and level of memory load. Along similar lines these authors also found that strong pseudohomophone associative priming, both with pseudohomophones as primes and as targets (Lukatela & Turvey, 1991). The author's conclude that their results suggest that lexical access is primarily phonological.

Alternatives to the dual-route model have been suggested. Van Orden also proposed an account in which phonological coding always mediates lexical access, although it is conceived within the framework of a connectionist system (Van Orden, 1987; Van Orden et al., 1991). Van Orden points out that there has been a bias among researchers to assume that direct access is a given while it is the role of phonology that is debated. In fact, he argues, it is possible to seriously question all of the existing data purporting to show direct access and consequently to treat direct access as the suspect construct. Other challenges to dual route theory have ranged from proposals of visually based access (e.g., Glushko's analogy theory), to attempts to create modified dual route accounts wherein GPC mapping occurs at several levels of structure, or GPC mapping is in some way sensitive to the statistical regularities in the mapping (Carr & Pollatsek, 1985). While dual route theory stands challenged in several ways, it still provides a useful framework within which to organize research questions, and the notion of more than one pathway to lexicon has not been made implausible by any of these results.

The experiments reported here were motivated by the idea that clear evidence suggesting a variable reliance on phonological or visual information would, among other things, obviously pose a challenge to any model that assumes a single route to lexicon. It might be possible to induce subjects to modify which type of coding they rely on in a word recognition task. Such evidence would not only be generally relevant to the study of reading, but would suggest a very fine degree of attentional control over relatively low level cognitive processes, and therefore would also be relevant to other areas of cognitive psychology. In the following section some data relevant to the question of processing flexibility is reviewed.

Evidence of flexible coding processes

As noted above, dual route theories usually assume that with increased reading skill or word familiarity reliance on orthographic information for accessing the lexicon should also increase. Such a developmental shift in reliance on type of code would constitute important support for dual route theory. Seidenberg and his colleagues (Seidenberg, Waters, Barnes, & Tannenhaus, 1984; Waters & Seidenberg, 1985; Waters, Seidenberg, & Bruck, 1984) found that regularity effects in the lexical decision task (faster response latencies to words that conform to GPC rules than to words that violate these rules), which would appear to implicate prelexical phonological processing, diminish both with increasing reading ability and with increasing word frequency within a given reading level. This has been taken to suggest a shift to reliance on direct access (but see Van Orden et al., 1990).

Evidence suggesting experimentally induced shifts in reliance on phonological information in several different types of word recognition tasks has been reported (Andrews, 1982; Hanson & Fowler, 1987; Hawkins et al., 1976; McQuade, 1981, 1983; Monsell et al., 1992; Paap & Noel, 1991; Shulman et al., 1978). In a study employing a two alternative forced choice recognition task with masked stimuli, Hawkins, Reicher, Rogers, and Peterson (1976) found that when the proportion of homophone pairs was high, and subjects were informed of this fact, they were no worse at choosing the correct target in homophone pairs than they were at choosing the correct target in non-homophone pairs. However, when the proportion of homophones was low subjects were significantly slower on homophone pairs than on non-homophone pairs. The authors argued that subjects were able to strategically control the extent to which they employed phonological coding.

Shulman, Hornak, and Sanders (1978) used a paired lexical decision task, wherein the subject decides if two letter strings, simultaneously presented, are both words. In one experiment, subjects received pronounceable nonwords and in a second experiment they received illegal nonwords. When subjects received pronounceable nonwords, latencies on pairs of words that were orthographically similar but phonologically dissimilar (e.g., COUCH - TOUCH) was inhibited relative to a control condition. However, when illegal nonwords were employed, performance on this type of pair was actually facilitated relative to control. They interpreted this as evidence that subjects in the former condition employed phonological codes while subjects in the latter condition did not. Hanson and Fowler (1987) essentially replicated this finding. However, Van Orden and his colleagues (Van Orden et al., 1990) have argued that this result can be interpreted within a phonologically oriented model if it is assumed that subjects in the illegal nonword context can rely on "noisy" as opposed to "cleaned-up" phonological codes. This issue is addressed in our third experiment.

Davelaar, Coltheart, Besner, and Jonasson (1978) conducted several experiments whose outcomes can be interpreted as suggesting strategic flexibility in reliance on phonological coding. In a lexical decision task, they manipulated whether or not the nonword context contained pseudohomophones (nonwords which, when pronounced, sound like real words; e.g., BRANE and BOTE). Words were either homophones (e.g., SALE) or matched nonhomophonic controls. In an initial condition with no pseudohomophones among the nonwords, low frequency homophonic words were responded to more slowly than their controls. However, in a second condition where the nonword context contained many pseudohomophones, no homophony effect was observed. These authors concluded that subjects can strategically control whether or not they use phonological coding.

McQuade (1981, 1983) also manipulated the proportion of pseudohomphones used a lexical decision experiment. One group of subjects received a high proportion of pseudohomophones, while a second group received a low proportion of these items. Performance on a common set of pseudohomophone targets was compared to performance on a set of matched nonword controls. Previous studies had shown that subjects tend to respond more slowly to pseudohomophones than to nonpseudohomophones; this has been referred to as the pseudohomophone effect. In the McQuade study the high-proportion pseudohomophone group showed no pseudohomophone effect whereas the low-proportion group did. McQuade surmised that the high-proportion group had suppressed phonological coding, since phonological codes would be misleading and disadvantageous on a large proportion of trials. Presumably, these subjects relied on visual access coding and, therefore, were not slower on the critical pseudohomophones than on the nonpseudohomophones. This finding, while suggestive, speaks primarily to nonword processing and does not necessarily provide insight into the processing of words.

Andrews (1982) also manipulated nonword context in a lexical decision experiment. Two groups of subjects received a common set of words, but for one group half of the nonwords were pseudohomophones, while for the second group no pseudohomophones were included among the nonwords. The pseudohomophone group was significantly faster on word trials than the nonpseudohomophone group. Andrews suggested that subjects in the pseudohomophone group bypassed the phonological route and, relying on the direct access route, were faster than the no-pseudohomophone subjects who were waiting for the output from the phonological route. However, a possible speed accuracy tradeoff was present in these data. Andrews also manipulated other characteristics of the words. She crossed regularity (regular vs. exception word) with consistency (absence or presence of neighbors with different rime pronunciations) and found that consistency was more reliably associated with latencies than was regularity. However, there were consistency effects for both groups and no interactions between the group variable and consistency were reported. On the view that subjects in the pseudohomophone group were, in some way, bypassing the phonological route, while the no pseudohomophone subjects were not, differences in the magnitude of phonological effects in the two conditions would have been expected, and this outcome was not obtained. The current Experiments 1 and 2 involve similar manipulations of nonword context and attempt to determine whether a pseudohomophone-induced speed difference coupled with a difference in the magnitude of phonological effects between the groups can be obtained.

In contrast to Andrews' results, Stone and Van Orden (1992) found a word response latency difference favoring a group that received no pseudohomophones over those who received 100% pseudohomophones in the nonword context of a lexical decision task (see James 1975 for similar results). This result directly opposes the one obtained by Andrews. Further, Stone (personal communication) reports that in some new as yet unpublished studies a latency disadvantage was observed for a group receiving only 50% pseudohomophones, and that would constitute a failure to replicate the outcome obtained by Andrews (1982). However, the differences obtained in these experiments are far from reconciled at this point (Stone & Van Orden, in press), and the current experiments provide further evidence along these lines.

Paap and Noel (1991) also manipulated context across groups using a naming task. One group of subjects was asked to pronounce a list composed exclusively of exception words, while a second group was given fifty percent exception words and fifty percent regular words. Performance on a common set of exception words was the variable of interest. Subjects who received all exception words were faster on the critical items than subjects in the mixed context. Paap and Noel argued that this finding is consistent with dual route theory. They claimed that because phonological coding is not efficient for exception words, subjects in the all exception word context bypassed assembled phonology and instead used addressed phonology to name the target. By relying on direct access, they processed words more quickly than subjects in the mixed list condition who, presumably, were engaged in a greater degree of assembled phonological coding. One problem for this interpretation lies in the fact that subjects in the all exception word context had, in effect, more naming practice with this kind of word and might have been faster on critical trials regardless of the route employed in lexical access. In the same study Paap and Noel also looked at naming performance under dual-task conditions (concurrent memory load task) and found that low frequency exception words were actually named *faster* under the high rather than under the low memory load condition. In contrast, low frequency regular words and both high frequency regular and exception words were all named more slowly under high load than under low load. They claimed that this effect came about because the assembled phonological route was more handicapped by concurrent attentional demands than the addressed phonological route; since the assembled phonological information is thought to primarily inhibit the naming of low frequency exception words, slowing it down through the use of a heavier memory load reduced its negative influence (but see Lukatela and Turvey in press for contrasting results using similar procedures).

Monsell, Patterson, Graham, Hughes, and Milroy (1992) also employed a naming task, and contrasted conditions in which lists consisted of words only, of nonwords only, or both words and nonwords. All words were of the exception type. They found that words presented in the word only list received fewer regularization errors than words presented in the mixed word/nonword context. They argued that since nonwords require the phonologic route in order to generate a pronunciation, subjects receiving the mixed word/nonword context relied more on this assembled phonology and hence made more regularization errors. Subjects receiving the exclusive word context, on the other hand, could rely on the lexically generated addressed phonology and therefore regularization errors were less likely. The authors proposed that subjects can strategically disable the assembled route in a naming task when conditions make it useful to do so.

Taken as a whole these studies seem consistent with the proposal that subjects are flexible in the degree to which they employ phonological codes in word recognition tasks. Further, these results have been obtained with several different word recognition tasks. The current experiments further explore the nature and consequences of coding flexibility. They begin with a quasi-replication of the basic phenomenon of coding flexibility together with a demonstration that the effect is indeed a phonological one: The effect is shown to involve the phonological similarity of the lexical neighborhood. In the second experiment, evidence is presented which suggests that the use of assembled phonological information makes measurable demands on attentional resources. Finally, the third experiment suggests that the extraction of meaning from an identified word is not affected by which route predominates in lexical access. This is consistent with the proposal that the effect of coding flexibility is on prelexical, not postlexical, processing.

A pilot study was conducted using a between-groups manipulation of pseudohomophony. One group of subjects in a lexical decision experiment received no pseudohomophones (NPsH group). The stimulus list for a second group was created by replacing 15% of the nonwords in the first list with pseudohomophones (PsH group). Both groups received identical word lists. Half of the 128 words were of low frequency and half were of high frequency. Results indicated that the word responses of subjects in the NPsH group were significantly slower than the responses of subjects in the PsH group (NPsH = 569 ms, PsH = 524 ms). Frequency was significant and there was a marginally significant interaction between group and frequency indicating that subjects in the NPsH group were more adversely effected by low frequency words than subjects in the PsH group. An analysis of the accuracy data revealed no significant differences between conditions. Hence, subjects who received pseudohomophones produced faster and no less accurate word responses than subjects who received no pseudohomophones. This outcome conforms to Andrews (1982), who used a much higher proportion of pseudohomophones (50%) but also found a latency advantage for subjects in the pseudohomophone condition, but conflicts with Stone and Van Orden's (in press) results in which a 100% pseudohomophone group was much slower than a NPsH group.

The latency advantage for subjects in the PsH condition does not appear to be attributable to a simple lowering of a response threshold criterion because that should result in a lower accuracy rate for this condition; subjects in this group were actually slightly *more* accurate (nonsignificantly) than subjects in the NPsH group. The between group differences obtained in the pilot study might be thought of as the consequence of the fact that subjects in the PsH group are in some way either disabling the phonologic route or, perhaps, are executing a response prior to its output.

A less interesting account of the results from this pilot study could argue that the speed advantage (without a corresponding increase in errors) comes about because subjects in the PsH group exert more cognitive effort (greater attention) due to the difficult homophony created by the pseudohomophone items. This attentional account would not require any assumptions about differences in type of coding between the two conditions. Experiment 1 was conducted to determine whether the observed between-group difference in latencies is also associated with differences in the magnitude of effects of phonological processing difficulty. That outcome would implicate processing type differences in the two conditions.

EXPERIMENT 1

A pseudohomophone manipulation was employed, as in the pilot experiment, in a lexical decision experiment. However, in Experiment 1 words were selected specifically to provide a broad range on two dimensions: frequency and phonological processing difficulty. Phonological processing difficulty was indexed for each target word by a count of the number of "unfriendly" neighbors, defined as the number of English words sharing the same orthographic rime (the same spelling) as the target word but differing in rime pronunciation (e.g., BOOT and FOOT). A target word's number of unfriendly neighbors can also be considered as an index of phonological inconsistency for that word. Some words contained no "unfriendly" neighbors while others contained many. This continuous measure of phonological processing difficulty is correlated with whether or not the word is regular or exceptional with regard to GPC rules (and many words of both types were contained in the list). However, there were several regular words with unfriendly neighbors and a few exception words with none. As noted above, the number of words that either share, or do not share rime pronunciation with the target would be psychologically important in any dual route theory where grapheme to phoneme mapping is sensitive to the statistical characteristics (such as frequency of occurrence) of these transforms (Rosson, 1985; Van Orden et al., 1990). By any such account, generating the appropriate GPC mapping for the target word will be more difficult as the number of words in the lexicon possessing the same orthographic structure but a different phonological realization increases. In any case, without theoretical commitment as to how consistency and regularity might differ, we noted that several lexical decision studies have found that indices of phonological processing complexity based on an examination of the target's phonological neighborhood are associated with performance (Andrews, 1982; Jared, McRae, & Seidenberg 1990; Perfetti & Bell, 1991). We predicted that subjects relying on phonological information during lexical access (NPsH condition) would be more sensitive to phonological processing difficulty than subjects engaged in direct access (PsH condition).

A recognition memory test was also conducted to determine whether subjects in these conditions differed in their depth of processing. For instance, if subjects in the PsH condition are faster as a consequence of failing to process the targets through to meaning while subjects in the NPsH condition are processing through to meaning, then episodic memory differences would be expected, since semantic processing is associated in recognition memory with superior performance (Craik & Lockhart, 1972).

Method

Subjects. Forty-nine undergraduate students from the College of the Holy Cross participated for partial course credit.

Stimuli. Two lists, each containing 128 monosyllabic words and 128 monosyllabic pronounceable nonwords were constructed. The only difference was that List 1 contained no pseudohomophones among the nonwords while List 2 contained 22 pseudohomophones (17%). Sixty nonhomophonic words were the critical experimental items chosen to provide a broad range of frequency (Kučera and Francis range = 2 - 1617) and phonological consistency values, and 68 words were fillers (included for use in a subsequent recognition memory test). Phonological inconsistency was indexed as the number of monosyllabic words that share the target's orthographic rime but which pronounce the rime differently than the target (range = 0 - 26). We called these words "unfriendly neighbors." The log of the number of unfriendly neighbors was used in the analysis, labeled simply NU. Length (number of letters) was included as a control variable (range = 3 - 6).

The filler words consisted of 30 homophones and 38 nonhomophones. After subjects finished the lexical decision task they were given a surprise recognition memory test. They were given a 140 word list and were asked to circle the words that they remembered form the lexical decision task (subjects were informed that half of the words on the list had appeared in the previous task). Included among the 70 previously viewed items were 15 of the homophonic filler words. Also included in the memory test were 15 words that subjects had not seen but which were homophonic to words used in the lexical decision task.

Procedure. Subjects were randomly assigned to either the pseudohomophone (PsH) group or the nonpseudohomophone (NPsH) group. A standard lexical decision procedure was followed. Items were presented in a different random order to each subject in uppercase letters on a Macintosh 512K computer screen. Targets were preceded with a 500 ms fixation point (asterisk) in the middle of the screen and a 500 ms blank. Target presentation continued until the subject's response or until 1600 ms had elapsed. Latencies shorter than 150 ms or longer than 1600 ms were recorded as errors. "Word" responses were made with the dominant hand and "Nonword" responses were made with the nondominant hand on two telegraph keys. RT was measured with an accuracy of ±2 ms. Subjects were given 40 practice trials. Following the 256 lexical decision trials subjects were given a surprise recognition test consisting of 140 words. They were informed that half of these words were from the lexical decision list and half were not, and were told to indicate the items that had been presented in the task by circling them. They were also instructed to work at a fairly quick pace. Each subject's participation lasted approximately thirty-five minutes.

Results

For each subject, mean latencies were calculated for the correct and incorrect word and nonword responses. Within each of these categories, trials with latencies greater than two standard deviations from the subject's own mean (calculated independently for each category) were treated as errors. Mean latencies were computed, averaging over subjects, for each experimental word and for each nonword that appeared in both the NPsH and PsH conditions. Accuracy for each of these items was calculated as the proportion of subjects responding correctly to it. One of the sixty experimental words failed to be displayed due a programming error and three other words had error rates greater than 60%; these three items were excluded from the latency analysis but not the accuracy analysis.

Standard multiple regression analyses were performed on word latency and accuracy with items as the unit of analysis. The following regressors were used: the log number of phonologically unfriendly neighbors (NU), log word frequency, the interaction between these two, word length, and pseudohomophone group (as a repeated measure). The interactions between the repeated measure and each of the other regressors were also included in the analyses, but were removed from an analysis if they were nonsignificant. This procedure was followed in all subsequent analyses. The categorical variable regular-exception and the *proportion* of neighbors that were unfriendly to the target (NU/ Total number of neighbors) was tried as well, but only NU was significantly associated with performance; hence all subsequent analyses use NU as the index of the phonological inconsistency of a target word's neighborhood.

For word latency there was a significant effect of group, $F(1, 54) = 9.96$, $MS_e = 424.25$, $p < .01$, indicating that mean latencies were faster in the PsH condition (513 ms) than in the NPsH condition (535 ms). There was a significant effect of NU, $F(1, 51) = 4.08$, $MS_e = 1706.07$, $p < .05$, and its positive regression coefficient (31.70) indicates that latencies increased with the number of unfriendly neighbors. While the main effect of frequency was not significant in this model, it should be noted that with the term representing the interaction between frequency and NU removed from the model frequency was significant, $F(1, 52) = 15.78$, $MSe = 1830.27$, $p < .001$. The interaction of NU with frequency was, however, significant, $F(1, 51) = 4.79$, $MSe = 1706.07$, $p < .05$. In order to examine this interaction we split the words into two roughly evenly sized frequency categories: low (including items with frequencies between 1 and 20) and high (including items with frequencies between 22 and 1617). The positive coefficient for NU, reliable in the overall analysis, was present only for low frequency words (14.04); the high frequency coefficient was actually negative (-2.73). Thus, the inhibitory effect of NU appears to have been largely carried by the lower frequency words. The interaction between group and NU was also significant, $F(1, 51) = 4.25$, $MSe = 405.22$, $p < .05$. Given the interaction between group and NU, data from the two groups were analyzed separately.

Table 1 summarizes the separate latency analyses of the data from the PsH and NPsH groups. Of critical interest is the fact that NU, as well as the interaction between NU and frequency, was significant *only* for the NPsH group (for NU: $F(1, 51) = 5.98$, $MSe = 1304.56$, $p < .05$; for the interaction: $F(1, 51) = 5.75$, $MSe = 1304.56$, $p < .05$). The positive regression coefficient for the NU effect indicates that latencies increased as the number of unfriendly phonological neighbors increased. The interaction revealed, as in the omnibus analysis, that this was especially true for the low frequency items (regression coefficients were: low freq. = 25.55, high freq. = -.607). In the PsH condition neither of these terms was significant.

The omnibus analysis of the accuracy data revealed that only NU was significant, $F(1, 54) = 5.02$, $MS_e = .033$, $p < .05$; as the number of unfriendly neighbors increased accuracy decreased (the coefficient for NU = -.145). Note that as with the latency data, without the frequency x NU interaction term in the model frequency was significant, $F(1, 55) = 11.93$, $MS_e = .035$, $p < .001$. There was no group difference (NPH = 87.2%, PH = 87.4%). However, in keeping with the latency analysis, the accuracy data from the two groups were also separately examined. Table 1 also summarizes the word accuracy data for the two groups. As with the latency analysis, NU and the NU by frequency interaction were significant for the NPsH group (for NU: $F(1, 54) = 6.82$, $MS_e = .016$, $p < .05$; for the interaction: $F(1, 54) = 5.30$, $MS_e = .016$, $p < .05$), but not for the PsH group. The negative regression coefficient for the NU effect indicates that as the number of unfriendly neighbors increased, accuracy decreased. As with the latency data, the words were divided into lower and higher frequency sets in order to examine the NU by frequency interaction. For lower frequency words the coefficient was negative (-.143), while for higher frequency words the slope was actually slightly positive (.016). Thus the accuracy results parallel the latency results in this regard.

Table 1. *Experiment 1: Analyses by condition.*

		Pseudohomophone Group		Non-Pseudohomophone Group	
		Coefficient	*F*	Coefficient	*F*
Latency:		$R^2 = .24$	$MS_{res} = 807$	$R^2 = .30$	$MS_{res} = 1305$
	NU	15.93	1.09	47.46	5.98 *
	Frequency	-9.73	2.45	-9.47	1.44
	NU * Frequency	-13.11	2.10	-27.57	5.75 *
	Length	—	< 1.00	—	< 1.00
Accuracy:		$R^2 = .18$	$MS_{res} = .02$	$R^2 = .28$	$MS_{res} = .02$
	NU	-.12	2.89	-.17	6.82 *
	Frequency	.04	1.70	.03	1.23
	NU * Frequency	.06	1.71	.09	5.30 *
	Length	—	< 1.00	—	< 1.00

* $p < .05$; ** $p < .01$; *** $p < .001$

An analysis was also conducted on the 106 nonwords that subjects in both conditions had received (not including the pseudohomophones or corresponding nonpseudohomophones that were unique to one or the other condition). As with words, correct rejection latencies were faster in the PsH condition (mean = 569 ms) than in the NPsH condition (mean = 596 ms), $F(1, 105) = 61.59$, $MS_e = 629.08$, $p < .001$. No significant accuracy difference was obtained.

Memory Results

The memory data for the recognition test that was administered after the lexical decision trials were analyzed using signal detection analysis. Mean d' for the NPsH group was 2.56 and was 2.57 for the PsH group. This difference was not significant ($F < 1.0$). An analysis of performance on just the homophonic targets and the foils also failed to reveal a significant difference between the NPsH and PsH groups.

Discussion

Experiment 1 revealed a latency advantage for words and nonwords in the pseudohomophone (PsH) condition over the nonpseudohomophone (NPsH) condition, even though the inclusion of pseudohomophones in the nonword context might have made the former condition more difficult, not less. This result replicates the results of the pilot study as well as the results reported by Andrews (1982) who used 50% pseudohomophones in the PsH condition. In the present experiment the latency advantage for the PsH group cannot be attributed to a speed-accuracy tradeoff since the PsH group was slightly more accurate than the NPsH group (although the difference between the two groups was not statistically significant). However, the latency disadvantage for a group receiving 100% pseudohomophones reported by Stone and Van Orden (in press) stands in contrast to the current results. Further, these authors argue that eliminating a slower route will not necessarily speed latencies, especially if a horse race process is assumed (Paap & Noel, 1991). The current results might be taken to indicate either that disabling the phonological route frees attentional resources thereby producing more efficient orthographic processing, or alternatively, that subjects who do not disable the phonological route in some sense wait for this information to build up and that this produces the longer latencies observed in the NPsH context. While the conflicts within the literature are not resolved, this experiment establishes, for the first time, a link between faster responding and diminished phonological influences.

There was additional evidence that phonology had been used for word recognition in the NPsH group but not in the PsH group (or, at least, not to the same degree). In the NPsH condition, the difficulty of phonological processing, as indexed by the number of phonologically unfriendly neighbors, had significant adverse effects on latencies and accuracy, whereas in the faster PsH condition this variable had no influence. This finding lends support to the claim made by Andrews (1982) that in a pseudohomophonic context subjects strategically inhibit phonological processing (because it tends to generate false positives), thereby shifting resources to the faster direct route. This is consistent with the basic architecture of dual route theories but is problematic for most single route models. Connectionist accounts (Seidenberg & McClelland, 1989; Lukatela & Turvey, 1990, in press; Van Orden et al., 1990) might cope with the results of Experiment 1 by recourse to a short-term, context induced, adjustment of network dynamics; however, only actual simulations can inform such speculation.

Even for dual route theories, the precise locus of strategic flexibility is not clear. It is possible that subjects disable or attenuate GPC level processing (Monsell et al., 1992). It is also possible that the strategic change occurs late; that subjects simply ignore the output from GPC level processing. It does seem unlikely that the flexibility demonstrated in Experiment 1 occurs at an even later post-lexical checking stage. If subjects in the NPsH group tended to engage in an extra step of "sounding out" an already recognized lexical representation then effects of the number of phonologically inconsistent neighbors should either not be of any consequence to this group or, alternatively, should affect both groups equally. Further, it is unlikely that the small latency advantage for the PsH group would be attributable to the elimination of what should be a relatively time-demanding postlexical check. It also seems that if the subjects in the NPsH group had adopted a more stringent criterion than the PsH group in checking the response, they would have had higher accuracy rates than subjects in the PsH group—but they did not. As noted earlier, Monsell and his associates (1992) and Paap and Noel (1991) both claim that subjects can ignore or bypass assembled phonological information when it is advantageous to do so in a naming task; the current results suggest a similar flexibility in lexical decision.

Finally, the recognition memory test was included to explore the possibility that subjects in the PsH group were obtaining a speed advantage by initiating their responses before processing the targets fully. Several recent lexical decision studies have suggested that subjects can use an orthographic familiarity bias, wherein a lexical decision is made without fully discriminating the target from its active neighbors (Johnson & Pugh, in press; Pugh, Rexer, Peter, & Katz, in press). It seemed reasonable that if subjects in the NPsH condition were more fully processing the target than subjects in the PsH condition (perhaps using target semantic information in making the decision), then episodic memory for the lexical decision stimuli would be superior in the former group. Had group differences been obtained the results would have been provocative; the failure to show group differences, however, should not be considered overly informative.

EXPERIMENT 2

The explanation that subjects in the pseudohomophone (PsH) condition were simply more *attentionally* focused than subjects in the nonpseudohomophone (NPsH) condition, as a consequence of the difficult pseudohomophone context, seems at odds with the fact that different neighborhood phonological inconsistency effects were also found for the two groups. The latter suggests a difference in the kind of processing, not simply a difference in the efficiency of processing. Nonetheless, such an account cannot necessarily be ruled out because greater attentional effort, if its influence reached down to the level of lexical processing, might suppress phonological ambiguity effects to some extent; of course attentional consequences of this sort would have relevance to the dynamics of word recognition. In order to examine the attentional characteristics of performance in the NPsH and PsH conditions, Experiment 2 employed a dual-task paradigm (Lukatela & Turvey, in press; Paap & Noel, 1991; Posner & Boies, 1971). As noted in the introduction, Paap and Noel (1991) found that naming latencies were influenced by the difficulty of a concurrent memory task. A similar manipulation was employed in the current experiment. Subjects were required to make lexical decisions while holding either one digit (low load) or four digits (high load) in short-term memory. Immediately after making the lexical decision, a target probe digit appeared on the screen and subjects decided whether the probe matched, or did not match, an element in the memory set. This memory load manipulation allowed us to examine lexical decision performance while attentional demands on the subject were either low or high.

If subjects in a PsH condition simply exert greater attentional effort in lexical decision, then increasing attentional resource demands by increasing memory load should cut into the available resource more for these subjects than for subjects in an NPsH condition. However, if the pattern of results obtained in Experiment 1 resulted from a selective disabling of the phonologic route because of the pseudohomophone context, then PsH subjects in Experiment 2, unencumbered by the presumably greater attentional demands of phonological processing, might not only perform the lexical decision task more efficiently, but might also perform the secondary memory task more efficiently.

Method

Subjects. Thirty undergraduate students from the University of Connecticut, participated for partial fulfillment of a course requirement.

Stimuli. Ninety-six experimental words, possessing a broad range of values on the dimensions of frequency (1 - 1617) and number of unfriendly neighbors (0 - 26), were used. Along with these, 36 filler words (32 of them homophones) were included, as in Experiment 1. The nonwords for the NPsH group were 132 pronounceable nonpseudohomophones; in the PsH condition, 30% of these nonwords were replaced with pseudohomophones. All of the word and nonword stimuli are presented in Appendix A. Half of the memory sets consisted of one digit (low load) and half consisted of four digits (high load). Within each of these, half of the target probes matched an element in the set, while for half there was no match. Each stimulus was viewed by half of the subjects under the one digit memory load and by the other subjects under the four digit memory load.

Procedure. Each trial began with the presentation of a fixation point (an asterisk) for 500 ms. After a 500 ms pause the memory set was presented for 1500 ms. Following a 1500 ms interval, the lexical decision target was then presented. "Word" responses were made with the dominant hand and "Nonword" responses were made with the nondominant hand on two telegraph keys. Subjects had 1500 ms from the onset of the stimulus to make their lexical decisions. 1700 ms after the offset of the lexical decision target the probe digit appeared on the screen for 600 ms and subjects had up to 1500 ms to respond positively or negatively using the same telegraph keys. Lexical decision and probe response latencies shorter than 150 ms or longer than 1500 ms were recorded as errors. Latencies were measured with an accuracy of ±2 ms. Subjects received forty practice trials and then the experiment's 264 trials, which were presented in a different random order to each subject. Subjects were instructed to respond as quickly and as accurately as possible to both the lexical decision and the memory probe judgment and were told that the word/nonword task was inserted into the retention interval in order to make the memory task more challenging. However, subjects were not explicitly told that the memory task was the primary task. Each subject's participation lasted approximately forty-five minutes.

Results

Mean lexical decision and memory probe response latencies were computed for each item, averaging over subjects following the same procedure that was used in Experiment 1. Accuracy values were also calculated as in Experiment 1.

Lexical Decision Analyses

Standard (simultaneous) multiple regression analyses were conducted on the latency and accuracy of the lexical decision word and nonword responses, using items as the unit of analysis, as in Experiment 1. The between-item variables in the analyses were log frequency (Frequency), log number of unfriendly neighbors (NU), the interaction between these two terms, and word length (Length). Group (NPsH vs. PsH) and Load (low vs. high) were within-item variables. Separate regression analyses were also performed on the NPsH data and the PsH data, as in Experiment 1.

Words: Latency. The omnibus analysis of the word latency data revealed a significant Group by Length interaction, $F(1, 91) = 8.65$, $MS_e = 880.65$, $p < .01$. As Figure 1 illustrates, the latency advantage for the PsH over the NPsH condition is larger for longer words (5 - 6 letters) than for shorter words (3 - 4 letters). The only other term to reach significance in this analysis was Frequency, $F(1, 91) = 17.06$, $MS_e = 5971.12$, $p < .001$; latencies decreased with increased frequency. It should be noted that with the interaction between group and length removed from the model the main effect of group did obtain significance, $F(1, 95) = 112.52$, $MS_e = 947.69$, $p < .0001$; NPsH (633 ms) and PsH (599 ms). The number of unfriendly neighbors (NU) did not reliably effect response latency nor did it interact with any of the other variables. The separate regressions performed on each Group revealed no additional effects.

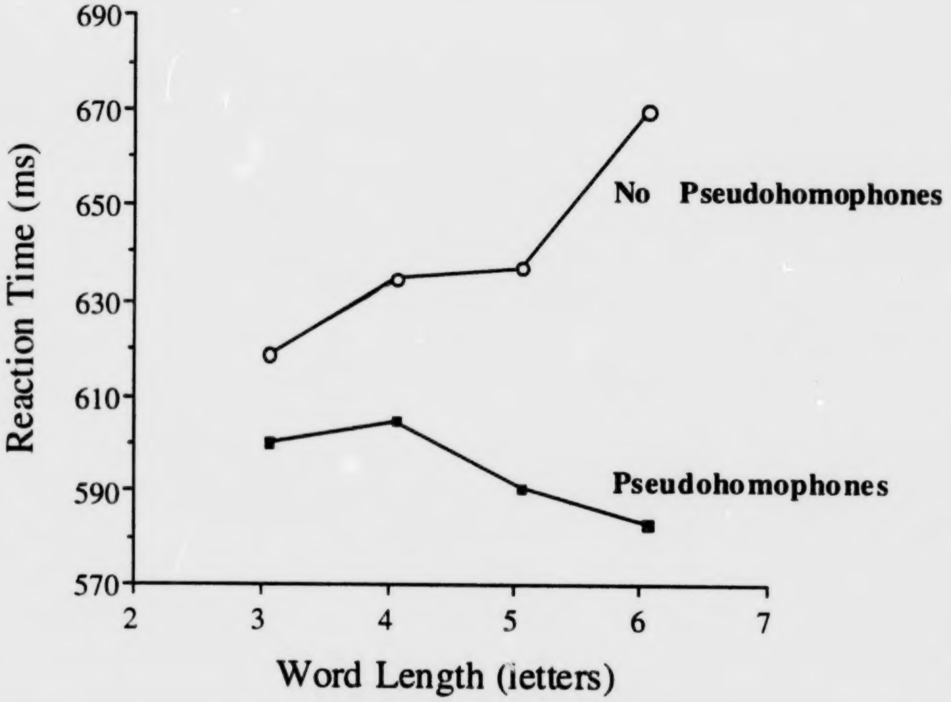


Figure 1. Word response latency in the dual-task procedure as a function of Group and word length.

Words: Accuracy. Neither the omnibus nor the separate regression analyses of the word accuracy data revealed any significant effects. The range of mean percent correct for the two groups across the two load conditions was only 2% (88%-90%).

Nonwords: Latency. The omnibus analysis of the latency data from the subset of nonwords that had been presented to subjects in both groups revealed a significant effect of Group, $F(1, 90) = 132.64$, $MS_e = 861.01$, $p < .001$; correct rejection latencies were faster in the PsH condition (674 ms) than in the NPsH condition (709 ms). There was also a significant effect of Load (low = 685 ms, high = 698 ms), $F(1, 90) = 14.15$, $MS_e = 1100.89$, $p < .001$; but no significant interaction between these two factors. Length was significant, $F(1, 90) = 3.98$, $MS_e = 5351.46$, $p < .05$, with longer nonwords yielding slower rejection latencies. A Load by Length interaction was also obtained, $F(1, 90) = 8.14$, $MS_e = 1100.89$, $p < .01$; the means were 672, 698, 695, and 701 ms for low (load)-shorter (length), low-longer, high-shorter, and high-longer conditions, respectively. Thus, the length effect was considerably larger under a low memory load (26 ms) than under high load (6 ms). Finally, the three way interaction between Group, Load, and Length approached significance, $F(1, 90) = 3.15$, $MS_e = 1809.11$, $p < .08$. The length effect for the NPsH group under low load was nearly twice as large as in any other cell (35 ms).

Nonwords: Accuracy. The omnibus analysis of the accuracy data revealed a significant effect of Group, $F(1, 90) = 10.21$, $MS_e = .0072$, $p < .01$, indicating greater accuracy in the PsH condition (92%) than in the NPsH condition (89%). No other terms were significant.

Probe Task Analyses

Standard multiple regression analyses were also conducted on the latency and accuracy of the probe task responses following words and nonwords, using items as the unit of analysis. These analyses used the same variables as the analyses of the lexical decision responses, and included the additional within-item variable of match or mismatch between probe and memory set (Match), and its interactions with the other variables. Separate analyses were also again performed on the NPsH data and the PsH data.

Words: Probe Latency. The omnibus analysis of the latency of probe responses following words revealed main effects of Load and Match, $F(1, 180) = 119.11$, $MS_e = 2632.35$, $p < .001$, $F(1, 180) = 38.37$, $MS_e = 2632.35$, $p < .001$, respectively. The interaction between these terms was also significant, $F(1, 180) = 12.34$, $MS_e = 2632.35$, $p < .001$. The means were 525, 433, 643, and 587 ms for the low-mismatch, low-match, high-mismatch, and high-match conditions, respectively. Thus, the interaction indicates that the advantage on match trials was larger under low load (92 ms difference) than under high load (55 ms difference). Group (NPsH vs. PsH) was also significant, $F(1, 190) = 87.90$, $MS_e = 2000.96$, $p < .001$, revealing that probe task performance was faster in the PsH condition (526 ms) than in the NPsH condition (569 ms). The interaction between Group and Load was also significant, $F(1, 190) = 6.22$, $MS_e = 2000.96$, $p < .05$. The means were 495, 642, 464, and 588 ms for the NPsH-low, NPsH-high, PsH-low and PsH-high conditions, respectively. Thus, the latency advantage of low load over high load was 23 ms larger under in the NPsH condition than in the PsH condition.

In the separate analyses of the PsH and NPsH groups the effect of NU was marginally significant in the NPsH condition, $F(1, 184) = 2.92$, $MS_e = 2463.52$, $p < .10$. The regression coefficient for NU was positive (18.36), thus as the number of phonologically unfriendly neighbors increased latencies increased. Also, the interaction between NU and Frequency was marginally significant in the NPsH condition, $F(1, 184) = 3.80$, $MS_e = 2463.52$, $p = .05$. As in Experiment 1, the effect of NU in the NPsH condition was examined separately for high and low frequency words. As expected, the regression coefficient for NU was positive for the lower frequency items (2.90), while for higher frequency items the coefficient was negative (-26.67). Neither of these terms approached significance in the PsH condition.

Words: Probe Accuracy. The omnibus analysis of the accuracy of probe responses following words also revealed a main effect of Load, $F(1, 180) = 8.78$, $MS_e = .004$, $p < .01$; the means were 97% and 94% correct for the low and high load conditions, respectively. There was also a significant interaction between Load and Match, $F(1, 180) = 7.10$, $MS_e = .004$, $p < .01$. The means were 97%, 96%, 95% and 92% for the low-mismatch, low-match, high-mismatch, and high-match conditions, respectively. The interaction appears to come from the fact there were relatively more errors on High Load trials when the probe matched an item in the memory set than in any of the other conditions. Group was also significant, $F(1, 191) = 7.10$, $MS_e = .004$, $p < .01$, indicating somewhat greater accuracy on the probe task for subjects in the PsH condition (96%) than for subjects in the NPsH condition (94%). No other effects or interactions were reliable, although NU approached significance, $F(1, 180) = 2.81$, $MSe = .004$, $p < .10$. An examination of the separate NPsH and PsH analyses revealed that NU was significantly related to accuracy only in the NPsH analysis [NPsH: $F(1, 180) = 4.10$, $MSe = .005$, $p < .05$; PsH: $F < 1.0$], and its negative regression coefficient (-.869) indicated that as the number of phonologically unfriendly neighbors increased, accuracy decreased. As with latencies, subjects in the PsH condition were not influenced by this variable.

Nonwords: Probe Latency. The omnibus analysis of the latency of probe responses following nonwords revealed main effects of Load, $F(1, 154) = 527.74$, $MSe = 2441.11$, $p < .001$, and Match, $F(1, 154) = 200.23$, $MSe = 2441.11$, $p < .001$. As with word trials latency advantages for both low load and match trials were obtained. The interaction between these terms was also significant, $F(1, 154) = 9.55$, $MSe = 2441.11$, $p < .01$. The interaction reflects the fact that the match advantage was larger (96 ms) under low load than high load (61 ms; the means were 565, 469, 675, and 614 ms for the low-mismatch, low-match, high-mismatch, and high-match conditions, respectively). Group was, once again, significant, $F(1, 158) = 60.29$, $MSe = 3234.60$, $p < .001$, revealing that performance on the probe task was faster in the PsH group (556 ms) than the NPsH group (605 ms). None of the interactions with group were reliable.

Nonwords: Probe Accuracy. The omnibus analysis of the accuracy of probe responses following nonwords indicated a main effect of Load, $F(1, 154) = 6.71$, $MSe = .004$, $p < .05$; accuracy was greater in the low load condition (95%) than in the high load condition (93%). Group was also significant, $F(1, 155) = 13.71$, $MSe = .004$, $p < .001$, with greater accuracy in the PsH group (95%) than the NPsH group (93%). No interactions were reliable.

Discussion

The main focus of Experiment 2 was to determine if lexical access made greater demands on attention for the NPsH group than for the PsH group (the latter condition is hypothesized to be less dependent on phonology). The results supported the hypothesis but in a manner that was less direct than expected. The memory load affected the NPsH group adversely (as expected) but its major effect was on that group's memory probe recognition rather than on its lexical decision performance. Responses to the memory probe were slower in the NPsH group and responses were less accurate. Further, for the NPsH group, increasing the memory load (from one to four digits held in memory) had a relatively greater deleterious effect on probe RT than for the PsH group. Thus, it was clear that subjects in the NPsH group were less able to perform the attention-demanding memory probe recognition.

Although the results of the lexical decision analyses were less straightforward, subjects in the NPsH group showed deficits in their performance on this task as well. The clearest results were for nonword lexical decisions, which were slower and less accurate for the NPsH group. With regard to lexical decisions on words, the only evidence of a NPsH-PsH group difference was an interaction between Group and word Length (see Figure 1); longer words were processed more slowly by the NPsH group but the reverse was true for the PsH group. This interaction suggests differences in the kind of processing and not simply in the efficacy or efficiency of processing. Any explanation that does *not* posit a change in type of processing would find this disordinal interaction problematic. To account for the interaction, we speculate that if the NPsH subjects (who are hypothesized to be relatively dependent on phonological coding) were engaged in a series of grapheme to phoneme conversions while processing a word, then a longer word should require more conversions and, therefore, should take longer to process. PsH subjects, on the other hand, apparently engaged in visual (i.e., orthographic) processing; perhaps this visual processing is a parallel rather than a serial process. Under parallel convergence, longer words would be processed faster than shorter words because longer words have smaller numbers of competing items in their respective neighborhoods (longer words are more nearly unique).

As noted above, the interaction between Group and Load in the probe recognition shows that subjects in the NPsH condition were more adversely affected by high load than PsH subjects. This is consistent with the idea that, due to extra phonologic processing in the NPsH condition, fewer attentional resources were available for the demanding memory condition. This interpretation is reinforced by the fact that as the neighborhood's phonological inconsistency (NU) increased, subjects in the NPsH condition were more likely to forget what they were holding in short-term memory (accuracy decreased). Additionally, as NU for low frequency words increased, NPsH subjects were slower on probe judgments. PsH subjects, on the other hand, who were hypothesized not to be dependent on phonological coding, showed no hint of an influence of neighborhood phonological inconsistency on probe recognition. In sum, the results suggest that subjects under pseudohomophonic (PsH) conditions adjusted processing in some way so as to eliminate the disadvantageous influence of phonological processing.

The fact that subjects in the PsH condition were generally faster and more accurate on *both* tasks than subjects in the NPsH condition seems consistent with the idea that subjects in the former condition were processing words in a way that was not only more advantageous for lexical decision, but actually freed up attentional resources for the memory task in which the word recognition task was embedded. If subjects in the PsH condition were in some way disabling the presumably attentionally expensive assembled phonology route, then the results can be explained. An alternative account which preserves the notion that GPC processes are in some way disabled in the PsH context, but does not borrow on the attentional resources explanation, might also be consistent with these data. Some residual interference between word recognition and memory probe tasks might result if the two share common code types (the memory set is likely held in STM by a phonological code). If so, then in the current experiment, subjects in the PsH condition, by not generating phonological codes, would suffer less interference than subjects in the NPsH condition (who engage in extra phonological processing during word recognition). In any event, the idea that the speed and accuracy advantages for PsH subjects in the first experiment could have come about due to greater attentional effort in lexical decision would seem to have predicted a trade-off between the two performances. Instead superior performance as a function of the inclusion of pseudohomophones was found on both tasks. An attentional account might be constructed that can handle aspects of the current results, possibly one assuming a great deal more vigilance in the PsH context, but this approach would seem less consistent with the total pattern of data than the coding flexibility hypothesis.

A somewhat perplexing aspect of the current experiment is that, while neighborhood phonological inconsistency only influenced the performance of NPsH subjects, as in Experiment 1, in this experiment the influence revealed itself, not on the lexical decision, but on the subsequent memory probe. How is it that subjects in the NPsH condition, if engaged in more phonologically based reading, would not exhibit NU effects on the word recognition task itself, as well as on the subsequent judgment? The failure to obtain a stronger NU effect for the NPsH subjects on lexical decision trials in this experiment cannot likely be attributed to a lack of statistical power; in fact, the non-significant regression coefficients were 11.01 and 29.37 for the NPsH and PsH groups, respectively, and this actually reverses the pattern observed in Experiment 1.

One possible account of these results is predicated on the following set of assumptions. Subjects in the PsH condition might have actively disabled the phonologic route since it signals false positive responses to pseudohomophones, and then relied exclusively on the direct route in making lexical decisions. This not only optimized lexical decision performance, it also freed attentional resources for the memory task and hence performance was enhanced there as well. Subjects in the NPsH condition did not disable the phonologic route, and in the first experiment, waited for the build-up of phonological information, and used it in making lexical decisions (hence the influence of NU). In Experiment 2, on the other hand, they also retained the phonologic route, but under the demanding conditions of the memory load context, they tended to make the lexical decision before the phonological representation was fully generated (hence no influence of NU and only a small unreliable latency disadvantage). Thus, they actually relied on the direct route to read out the lexical decision response. Nonetheless, since the phonologic route was not actively suppressed by these NPsH subjects, it continued to operate, and for words with many unfriendly neighbors it was particularly resource demanding. This unsuppressed activity might then impair the subsequent memory judgment performance; thus as NU increased accuracy on the probe task actually decreased. This speculation is ultimately grounded in the view that phonologic information builds up more slowly than direct access (but see Stone and Van Orden, in press, for a contrasting view), and that even subjects in the NPsH condition can make a word/nonword decision prior to completion of the phonologic process. However, pseudohomophones in the nonword context drive subjects not simply to make a decision prior to the completion of phonological processing but, instead, to actively suppress it. Obviously, such an account is speculative and is contingent on the view that strategic control operates at several points early in processing. Nevertheless, it seems to provide the only account that can handle the results of the two experiments. The possibility that subjects can either ignore phonological information or disable it, should be investigated further. Still, it remains the case that even in this experiment when NU had an influence on performance (in the probe task) it was only for subjects in the slower NPsH group, and once again The faster PsH subjects showed no sensitivity to this variable.

EXPERIMENT 3

The results of Experiments 1 and 2 are consistent with an account of lexical access that allows for strategic control over the extent to which phonological coding mediates lexical access. It is clear that there is a general performance advantage for subjects in the pseudohomophone (PsH) condition, coupled with the failure to observe any evidence that these subjects were sensitive to the phonological inconsistency of its target's neighborhood (i.e., the number of unfriendly neighbors, NU). However, the argument that the failure to observe effects of NU indicates the absence of phonological coding obviously hinges on the validity of NU as a diagnostic criterion. While there is evidence from Experiments 1 and 2 that NU effects are, in fact, useful, they are somewhat variable in magnitude (and, therefore, in reliability) between experiments. Thus, it is important to supplement the evidence provided by NU in order to provide converging evidence that the PsH and NPsH groups differ in their degree of phonological coding. To this end, we decided to seek converging evidence of phonological processing flexibility in a double lexical decision study.

The basic task involves presenting the subject with two letter strings (one above the other), with the subject responding positively if both are words and negatively if one or both are not. The relations between the two words in a given pair can be varied to measure orthographic, phonological, or semantic processing. In an initial investigation Meyer and his colleagues (Meyer, Schvaneveldt, & Ruddy, 1974) examined performance on words that were either orthographically and phonologically similar (e.g., BRIBE - TRIBE, LOOK - BOOK) or orthographically similar but phonologically dissimilar (e.g., COUCH - TOUCH, LEMON - DEMON). Relative to control conditions consisting of the same words in unrelated pairings (e.g. BRIBE - BOOK, TOUCH - DEMON) they found a small facilitatory effect for BRIBE - TRIBE pairs, and an inhibitory effect for COUCH - TOUCH pairs. They interpreted this result as evidence that subjects were employing phonological codes whereby the consistency of the pronunciations of the rime in BRIBE - TRIBE type pairs produced facilitatory priming, and the inconsistency of the two rime pronunciations in COUCH - TOUCH type pairs produced inhibitory priming. Shulman and his colleagues (Shulman, Hornak, & Sanders, 1978) replicated this finding when the nonwords were orthographically legal. However, when illegal nonwords were used (either consonant strings or random letter strings which violated orthotactic rules) they found facilitatory effects for both BRIBE - TRIBE and COUCH - TOUCH pairs. On the possibility that this result was obtained because subjects were making decisions at a prelexical level based, perhaps, on orthographic familiarity, they included an associatively related condition (OCEAN - WATER). Because they observed associative facilitation in both conditions, they suggested that subjects were activating lexical representations. They interpreted their results as suggesting that subjects in the illegal nonword condition were getting to lexicon without mandatory phonological coding. However, they did not compare the magnitude of semantic priming in the two conditions, since the nonword manipulation was across experiments. Hanson and Fowler (1987) replicated the Shulman et al. (1978) finding that with illegal nonwords facilitatory priming for COUCH - TOUCH pairs was obtained, and this held for both hearing and deaf readers. However, they did not include OCEAN - WATER type pairs in their study.

Recently Van Orden and his colleagues (Van Orden et al., 1990) noted that while facilitation for COUCH - TOUCH stimuli could be seen as one of the few positive findings supporting the existence of a direct route, it is subject to an alternative interpretation. They argued that when the nonwords are illegal subjects can rely on partial phonological information—"noisy" phonological codes—to recognize words, and since COUCH and TOUCH have a good deal of phonological overlap, they can still partially prime each other phonologically. When nonwords are legal and, therefore, the discrimination is more demanding, noisy coding is too error-prone and subjects rely on a "cleaned-up" phonological code; here the COUCH - TOUCH inhibition is found. In other words, they suggest that the differences are due to quantitative and not qualitative differences in processing.

The current experiment provides a test of Van Orden et al.'s hypothesis. Instead of using illegal nonwords to induce a shift away from phonological processing (as in previous experiments using this paradigm), a pseudohomophone manipulation was once again employed (as in Experiments 1 and 2).

All subjects received nonwords that were legal (thus, presumably difficult); the intention was for subjects in both groups to rely on cleaned up phonological codes and not to rely on orthographic representations. By this account, so far, one might predict COUCH - TOUCH inhibition in both the NPsH and PsH conditions. However, if subjects in the PsH condition disable or weaken the phonologic route (in spite of the difficulty of the nonwords) then COUCH - TOUCH *orthographic* facilitation should be obtained because there will be no basis for any phonological competition that would result in inhibition. In contrast, subjects in the NPsH group (who are presumably more dependent on phonological coding) should still show COUCH - TOUCH inhibition. Further, if subjects in the PsH condition have been gaining a speed and accuracy advantage in Experiments 1 and 2 by somehow making lexical decisions without really achieving lexical access (for example, by means of an orthographic word familiarity check), they might show COUCH - TOUCH orthographic facilitation but they would not be expected to show OCEAN - WATER semantic facilitation. Thus, the current experiment can converge with Experiments 1 and 2 to show that PsH subjects attenuated their phonological processing. It also provides a test of Van Orden's noisy-code account of the Shulman data and it tests for the possibility that subjects in the PsH condition are engaging in only shallow (non-semantic) processing. Further, in the first two experiments a number of the nonwords were orthographically unusual (small neighborhood items). In this experiment nonwords are, on average, more orthographically familiar patterns.

Method

Subjects. Forty-six undergraduate students at the College of the Holy Cross participated in the experiment for partial fulfillment of a course requirement.

Stimuli. Subjects received 96 word/word pairs (positive trials) and 90 word/nonword and 6 nonword/nonword pairs (96 negative trials). Six types of word/word pairs were prepared. Type 1 pairs were orthographically and phonologically similar (BRIBE - TRIBE, LOAD - TOAD). Type 2 pairs were controls that were orthographically and phonologically dissimilar. The control pairs were generated by pairing dissimilar Type 1 items (e.g., BRIBE - TOAD, LOAD - TRIBE). Type 3 pairs were orthographically similar and phonologically dissimilar pairs (COUCH - TOUCH, GONE - BONE). Type 4 pairs were controls for the Type 3 pairs, constructed in the same way as the Type 2 controls. These experimental pairs (Types 1 and 3) were the same as those used by Meyer et al. (1974) and Hanson and Fowler (1987), and Type 1 and Type 3 pairs were matched as closely as possible for length and frequency (see Meyer et al. for details). Type 5 pairs were semantically related pairs (OCEAN - WATER) chosen from the norms of Battig and Montague (1969), and Type 6 pairs were controls for the Type 5 pairs, again generated by rearranging the Type 5 pairs. Type 5 pairs were chosen from among the top five exemplars of each category in the norms; this was done to insure that each member of a related pair was a good category exemplar. However, this constraint did not allow for a matching of these pairs with Type 1 or Type 3 pairs on dimensions such as length and frequency. Thirty-two pairs of each word/word pair type were prepared (all stimulus pairs are presented in Appendix B). Two stimulus lists were constructed. List A consisted of 16 Type 1 pairs, 16 Type 3 pairs, and 16 Type 5 pairs, with the words from the remaining 16 pairs of each type rearranged to serve as controls (Types 2, 4, 6); in List B the situation was reversed. Thirty-two of the 96 negative trials consisted of orthographically similar items (e.g., LOOK - DOOK); this matched the number of positive pairs that were orthographically similar (Types 1 and 3). Thus, orthographic similarity was not correlated with whether the pair was a positive or negative response type. Half of the word/nonword pairs were presented with the word as the upper display item, and half with the word as the lower display item. Subjects in the nonpseudohomophone (NPsH) condition received either list A or list B as they are described above. Subjects in the pseudohomophone (PsH) condition received one of these lists with a pseudohomophone substituted for the nonword item in 30% of its word/nonword pairs. Both NPsH and PsH subjects received the same 32 orthographically similar negative trials.

Procedure. The procedure was the same as in Experiment 1 except stimulus pairs instead of single letter strings were presented (they appeared one above the other in the center of the screen), and subjects had 1500 ms to respond to the items. As in the other experiments, subjects received a practice list of 40 trials before the experimental trials. Each subject's participation lasted approximately twenty-five minutes.

Results

For each subject, mean latencies were calculated for the six types of word pairs and for correct responses to the two types of nonword pairs. Within each of these categories, trials with latencies greater than two standard deviations from the subject's own mean (calculated independently for each category) were treated as errors. The data for the subjects and items analyses were based on these data. The data from three subjects who made more than 30% errors in at least two response categories were excluded from further analyses.

Word Analyses. The primary analyses were conducted on the matched Type 1 and Type 3 pairs and their respective controls (Type 3 and Type 4). Type 5 and Type 6 pairs were not matched with the first four pair types (see stimuli section), and were examined separately, in order to determine whether there were group differences in the magnitude of semantic priming. Analyses of variance were conducted on the latency and accuracy data using both subjects (F_1) and items (F_2) as random factors. For the subjects analyses, mean latencies and proportions correct were computed for each subject for each of the six pair types. In the items analyses, mean latencies were computed for each of the experimental words, averaging over subjects, and accuracy was calculated as the proportion of subjects responding correctly to each item. In the subjects analyses, pair type was a within-subjects factor and group (NPsH vs. PsH) was between. These designations were reversed in the items analyses. List (A vs. B) served as an additional control variable in these analyses.

The latency analysis on the data from the first four pair types yielded a significant main effect of pair type, $F_1(3, 117) = 19.30$, $MSe = 6349.48$, $p < .001$, and $F_2(3, 120) = 8.98$, $MSe = 23811.03$, $p < .001$, and a significant interaction between pair type and group, $F_1(3, 117) = 2.93$, $MSe = 6349.48$, $p < .05$, and $F_2(3, 120) = 3.98$, $MSe = 6731.47$, $p < .01$. Separate analyses on the two pair types and their respective controls indicated no significant Group x Pair Type (experimental vs. control) interaction in the orthographically and phonologically similar condition. However, as expected, the interaction was significant in the orthographically similar but phonologically dissimilar condition, $F_1(1, 39) = 5.42$, $MSe = 7808.70$, $p < .05$, and $F_2(1, 60) = 6.04$, $MSe = 9231.01$, $p < .025$. (Orthographically and phonologically similar condition subject means were: NPsH experimental pair type = 798 ms, NPsH control pair type = 903 ms, PsH experimental pair type = 804 ms, and PsH control pair type = 912 ms. Orthographically similar but phonologically dissimilar condition subject means were: NPsH experimental pair type = 945, NPsH control pair type = 895, PsH experimental pair type = 875, and PsH control pair type = 912.) Figure 2 shows the differences in response latency between the experimental pair types and their respective controls. Positive numbers indicate that the experimental pairs were faster than their controls (facilitatory effects) and negative numbers indicate that experimental pairs were slower than controls (inhibitory effects). Subjects in both groups showed facilitation to orthographically and phonologically similar pairs. However, for orthographically similar but phonologically dissimilar pairs NPsH subjects showed inhibitory effects while PsH subjects were facilitated on these pairs relative to the control condition; hence the two way interaction noted above. There was also a significant List by Pair Type interaction, $F_1(3, 117) = 5.38$, $MSe = 634.48$, $p < .01$, and $F_2(3, 120) = 3.24$, $MSe = 23811.03$, $p < .05$. The cell means indicated that facilitatory effects for Type 1 pairs were larger for the B list than the A list, and that inhibitory effects on Type 3 pairs were smaller for the B list than the A list. However, the three-way interaction between List, Pair Type, and Group was not significant in either the subject or item analyses; the critical Group by Pair Type interaction was not qualified by List. Finally the 10 ms latency advantage for PsH subjects was not significant in either analysis.

The latency analysis on the data from the semantically related pairs (Type 5) and their controls (Type 6) revealed a main effect of pair type, $F_1(1, 39) = 61.11$, $MSe = 3156.72$, $p < .001$, and $F_2(1, 60) = 35.10$, $MSe = 8349.30$, $p < .001$. The semantically related pairs were responded to 94 ms faster than the control items (related mean = 718 ms, unrelated mean = 812 ms). Of critical interest, however, is that this variable did not interact with group in either the subject or item analyses (p values > .25). Figure 2 presents the differences between the experimental and the control latencies for the two groups: 105 ms and 85 ms for the NPsH and PsH groups, respectively. Thus, the magnitude of semantic priming effects was quite large for both groups.

The analyses conducted on the accuracy data showed no significant effects of Group, Pair Type, or their interaction.

Nonword Analyses. Analyses of the nonword data included the following variables: Group and similarity (orthographically similar vs. dissimilar pairs). An effect of group on latencies was significant in the item analysis, $F_2(1, 66) = 5.98$, $MSe = 2487.51$, $p < .05$, but not in the subject analysis ($F < 1.0$). Subjects in the NPsH condition (956 ms) were faster on correct rejections than PsH subjects (976 ms).

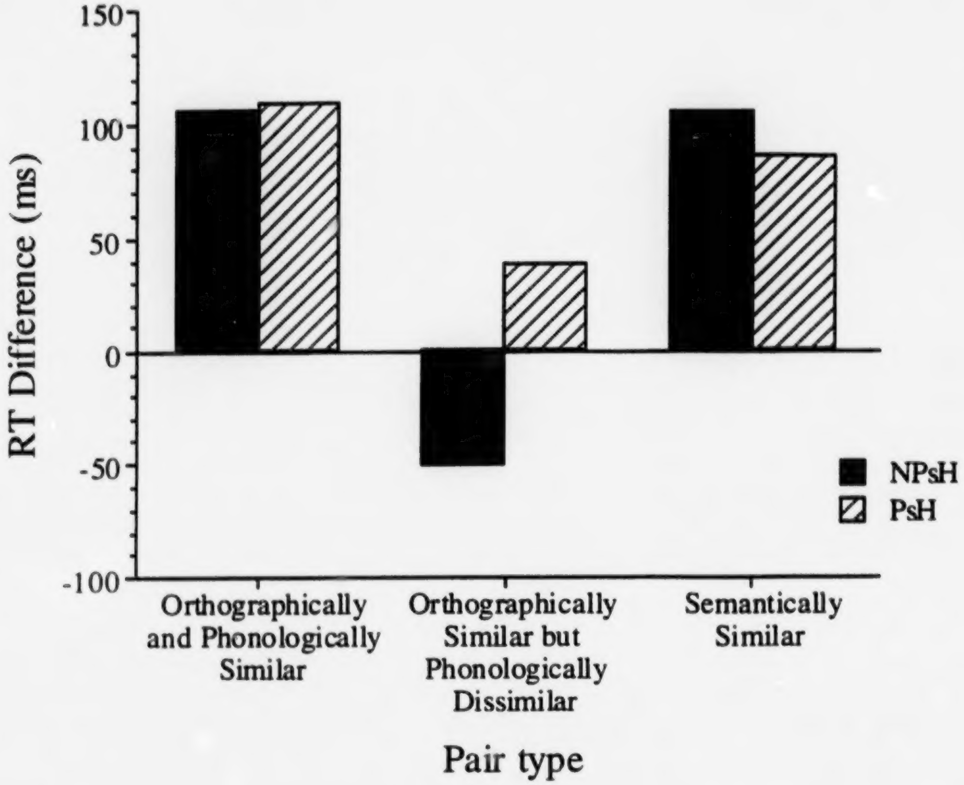


Figure 2. **The difference between experimental and control response latencies in the double lexical decision task as a function of Group and pair type. RT difference is the control group mean minus the experimental group mean.**

Discussion

The results of this experiment are quite clear with regard to the question of whether phonological processing differences between NPsH and PsH subjects are present. Subjects in the NPsH condition showed inhibitory effects when pair members were orthographically similar but phonologically dissimilar (COUCH - TOUCH) along with facilitatory effects when pair members were similar on both dimensions (BRIBE - TRIBE). In contrast, PsH subjects showed facilitatory effects of both types of pairs. The hypothesis that subjects in the PsH condition curtail phonological processing is strongly supported by these data.

Semantic association facilitated the responding of subjects in both groups. Note that while the magnitude of the effect was slightly greater in the NPsH condition, the difference was not reliable. Apparently, the subjects in the PsH group were able to get to lexicon without phonological representations of the target words. This was the same conclusion supported by Experiments 1 and 2, which showed no influence of the number of the target word's phonologically unfriendly neighbors in this pseudohomophone condition.

It should be noted that while this experiment revealed clear differences in phonological processing between the PsH and NPsH groups, the latency advantage for PsH subjects obtained in four other experiments (Andrews [1982], our pilot, and experiments 1 and 2) was not reliable in the current experiment. Perhaps when the judgment involves two letter strings the additional cognitive processing obscures the latency advantage that might be obtained due to attenuating or curtailing phonological processing. Further, for nonwords responses were actually faster in the NPsH condition although this was reliable only in the item analysis (subjects $F < 1.0$). This reverses the latency advantages in the PsH condition obtained in the other experiments on nonword trials. However, this weak effect favoring the NPsH subjects on nonword latencies should not be taken as evidence that somehow PsH subjects engaged in more phonological processing. If this had been so then the COUCH-TOUCH trial inhibition should have been observed for these subjects as well. Further, PsH subjects were not slower in rejecting pseudohomophones than regular nonwords (see the general discussion for a full discussion of nonword pseudohomophone effects across experiments). The cognitive differences between the single and double lexical decision merit further exploration.

The facilitatory effect of PsH subjects on COUCH - TOUCH pairs is not consistent with claims that the strategic effects that we have been documenting in this study are trivially postlexical in origin. If, for instance, phonological lexical access were mandatory, as several researchers have suggested (Lukatela & Turvey, 1990, in press; Van Orden et al., 1990), and subjects in the NPsH group were simply engaging in an extra postlexical phonological check, which the PsH subjects suppressed, then COUCH - TOUCH facilitation for PsH subjects would *not* be expected. Instead, by their account some obligatory inhibition on these pairs would be predicted, albeit of possibly smaller magnitude than in the NPsH condition. The observed facilitation suggests an orthographic based lexical access which is consistent with dual-route theory. Further, the claim that the PsH subjects in Experiments 1 and 2 were engaged in a kind of orthographic word familiarity judgment in lieu of actual lexical activation is not consistent with the large semantic priming effects observed in this experiment.

Following the hypothesis that PsH subjects disabled the phonologic route, BRIBE - TRIBE facilitation should be no different from COUCH - TOUCH facilitation for these subjects. However, that was not the case; for subjects in the PsH condition the magnitude of the facilitation for BRIBE - TRIBE pairs was more than twice as large as for COUCH - TOUCH pairs. A possible explanation for this difference is that most, but not all, subjects in the PsH condition showed facilitation on COUCH - TOUCH pairs, while nearly all showed BRIBE - TRIBE facilitation. Apparently, not all subjects responded to the pseudohomophone manipulation by disabling the phonologic route, although a significant proportion apparently did. These individual differences in response to contextual manipulations should be examined in subsequent investigations of strategic flexibility (see also Hanson & Fowler, 1987).

GENERAL DISCUSSION

The goal of this study was to demonstrate coding flexibility in lexical access. The three experiments varied the composition of the nonwords in a lexical decision task; in each experiment one group of subjects received pseudohomophones (PsH) among its nonwords and the other group did not (NPsH). The intention was to make dependence on phonological assembly counterproductive for the PsH group since the phonological realization of pseudohomophones falsely represents them as real words; this should lead to greater reliance on orthographic coding if this is, in fact, possible. In the first experiment subjects in the PsH condition performed faster and no less accurately than subjects in the NPsH condition on both word and nonword trials, suggesting that the presence of pseudohomophones had the predicted effect. Moreover, the performance of PsH subjects on word trials was uninfluenced by the phonological inconsistency of a target's orthographic neighborhood, while the latency and accuracy of NPsH subjects' responses were inhibited by neighborhood phonological inconsistency. Together these results suggest that subjects in the PsH group did not depend on assembled phonology to access lexicon (at least not to the extent of subjects in the NPsH group) but that they were, nevertheless, more efficient than the NPsH group.

In Experiment 2 similar latency and accuracy advantages for PsH subjects were found on the lexical decision and memory probe components of a dual-task procedure. However, unlike the outcome of Experiment 1, neighborhood phonological inconsistency did not affect lexical decisions in the NPsH group. Instead, inconsistency influenced performance on the memory probe that followed lexical decision for these NPsH subjects, with no corresponding influence on PsH subjects. Further, subjects in the NPsH condition were more adversely affected on probe performance by high memory load than were PsH subjects. These effects can be interpreted as suggesting that processing for the NPsH subjects not only involved phonological coding but was also more attentionally demanding. The interaction between group and word length on lexical decision latency suggested differences in the type of processing and not merely differences in efficacy or efficiency.

Experiment 3 used a double lexical decision paradigm to examine the influence of this phonological coding flexibility on phonological consistency effects. While subjects in both groups showed facilitation on phonologically consistent pairs (e.g., BRIBE - TRIBE), NPsH subjects showed inhibition on phonologically inconsistent pairs (e.g., COUCH - TOUCH). PsH subjects, however, were facilitated on these pairs, suggesting once again that they were not relying on phonological coding. Further, facilitation on semantically related pairs (e.g., OCEAN - WATER) was equivalent in both conditions, suggesting that subjects in both groups were, in fact, activating lexical entries. Apparently, dependence on direct access does not diminish activation of semantic information.

The clear implication from these studies is that, in the presence of pseudohomophones, a substantial proportion of subjects will process words so as to minimize phonological influences. The precise locus of this flexibility is unclear but there are several reasons to suppose it is not trivially postlexical. If it is not, then this poses a problem for single route theories in general, which would seem compelled to place coding flexibility—evidence for two kinds of processes—at some postlexical cognitive stage. As an often-proposed example of a postlexical mechanism, consider confirmatory postlexical phonological checking, performed after a word has already been selected in lexicon but before the response is made. The check uses a phonological representation of the target; if the representation does, indeed, "sound" identical to a word in the subject's speech lexicon, the original printed stimulus is confirmed to be a word. There are two possible sources for such a phonological representation: prelexical and lexical. It seems implausible that the former would ever be used when the latter is available; prelexical (i.e., assembled) phonology is typically incomplete—importantly, syllable stress for multisyllabic words is not indicated in the print and, therefore, cannot be present in the prelexical representation. Yet syllable stress is critical for the identification of spoken words. On the other hand, once a lexical entry has been activated (whether by assembled or direct processes), its complete phonological representation, including stress, is available. However, after lexical access, both conditions are identical with regard to access of lexical phonology. Thus, this assessment predicts (contrary to fact) that there will be no difference on words between conditions as a function of their postlexical processing because processing should be identical for both PsH and NPsH at that point. Nevertheless, there will be no such lexical phonology for pseudowords. Here, pseudohomophones will prove to be problematical for the PsH subjects. Suppose, therefore, that they completely eliminate the postlexical check. Because the NPsH subjects would not suppress the postlexical test, we would appear to have a possible explanation of the speed advantage of the PsH condition; the PsH subjects perform one less operation than the NPsH subjects (although we might expect the latency differences to be somewhat larger than they actually were). However, we should also see an elevated error rate for PsH subjects, because they are eliminating the check. In fact, the opposite result obtained; PsH subjects were slightly more accurate (nonsignificantly) even while performing faster. There are other difficulties encountered by an explanation based solely on post-lexical checking differences. If initial lexical access for both groups was phonologically mediated as suggested by several researchers (Lukatela & Turvey, 1992; Van Orden et al., 1990) then some mandatory inhibition on phonologically dissimilar pairs in the third experiment should have been seen in both groups, and that was not the case. If, on the other hand, pre-lexical processing was primarily orthographic for both groups, and phonological influences only occurred at a later stage, then it seems unlikely that NPsH subjects would shown increased sensitivity to phonological neighborhood inconsistency on the subsequent memory probe judgment and not on the initial lexical decision in the second experiment. Certainly, however, additional experiments are needed to examine the precise mechanisms involved in coding flexibility. At present however, it would appear that the most plausible account of the results of these three experiments is one that emphasizes context induced differences at the level of lexical access.

To the issue of whether subjects in the PsH condition either disabled the assembled phonological route or alternatively, simply read out responses prior to the completion of the phonological processing there are several findings which suggest the former interpretation. First, as noted above, since Experiment 2 probe task performance was influenced by phonological neighborhood inconsistency for NPsH subjects, suggesting a spill-over effect, if PsH subjects had merely been reading out quick responses and had not been disabling phonological processing a similar spill-over would still have been predicted in that condition. However, probe task performance was uninfluenced by NU for this group. Second, quick orthographically based responses should not be possible for nonword trials and consequently nonword performance should not have indicated group differences. In two of the three experiments PsH subjects were faster on nonword responses than NPsH subjects. To examine this issue further, we compared performance on regular nonwords and pseudohomophones for the PsH subjects. Note that these two sets of items were not specifically equated on any dimensions. However, one dimension that they do differ on is the phonological dimension and if subjects in this condition were processing nonwords in a phonologically sensitive fashion, then pseudohomophones' rejection latencies should have been somewhat slower than regular nonwords. In all three experiments rejection latencies were actually somewhat faster for pseudohomophones. The means for the regular nonwords and pseudohomophones were 569 vs. 566 ms in Experiment 1, 674 vs. 654 ms in Experiment 2, and 977 vs. 965 ms in Experiment 3. Thus, there was no hint of the standard pseudohomophone effect in any of these experiments. Such an outcome strongly implies that even on nonword trials subjects in the PsH group were operating in a non-phonological mode. There seems to be every indication in these data that PsH subjects were performing the lexical decision task in a fundamentally different way than NPsH subjects.

It should be noted that the interpretations being considered here are grounded in the idea that subjects are, in a strategic sense, altering the word recognition process. However, selective inclusion or elimination of pathways in lexical processing is not the only possible means of strategic control over performance. Stone and Van Orden (in press), in considering the results of their pseudohomophone context manipulation (see above), contrast pathway selection accounts with accounts based on flexible criterion setting . They find both accounts lacking in some ways but are more inclined toward the latter approach. A criterion setting account of the results of the experiments reported here does not appear to be very plausible. First, if subjects in the NPsH condition had simply set higher word and nonword response thresholds, resulting in longer latencies on word and nonword responses (Experiments 1 and 2), and this somehow amplified phonological influences, then correspondingly higher accuracy rates should have been observed in that condition,. As noted accuracy was slightly greater in the PsH group. Second, in Experiment 3 differential phonological sensitivity on COUCH-TOUCH trials was observed without any corresponding group differences in latency or accuracy. Again, the notion of differential use of phonological coding seems most plausible.

It might be tempting to conclude, given the fact that it took the "unnatural" presence of pseudohomophones to force subjects to adopt an apparently nonphonological mode of processing, that NPsH subjects are more representative of how normal word recognition operates. However, it is entirely possible that in a lexical decision task, in which half of the letter strings have no lexical representation, subjects occasionally adopt an inordinately phonological strategy. Given the plausibility of both arguments, it remains for experiments using more naturalistic reading tasks than lexical decision to resolve the issue of what constitutes normal phonological involvement for skilled readers (cf. Pollatsek, Lesch, Morris, & Rayner, 1992). We suggest, however, that the very existence of flexibility suggests that both phonological and direct processing are required in everyday reading, and that coding flexibility is, therefore, highly practiced. If not, why would the flexibility that we have demonstrated occur at all? If subjects always used only one strategy (at least since they became skilled readers) why should they be able to switch with apparent efficiency to another strategy (even if that switch is only partial, as from a single coding strategy to a mixed strategy)? In this regard, it is worth mentioning that our subjects may have had very little insight into the effect of the manipulation on their reading. Subjects appear to be exquisitely sensitive to the pseudohomophone manipulation, but nevertheless also appear to be unaware of what effect the pseudohomophones have on their process of word recognition. Anecdotally, we can report that when we queried a number of PsH subjects after the experiment, several were unaware that any of the nonwords were pseudohomophones. Those that were aware of the pseudohomophones claimed that they were forced to "slow down and be more careful." As we have seen, the opposite was the case: responses were faster in the PsH condition.

In conclusion, the current experiments suggest that subjects can control the extent to which they engage phonological processing in making lexical decisions. Further, in conditions where they apparently disable or attenuate such processing, lexical decision *and* concurrent STM based performance is enhanced. That this adjustment does not come at the expense of lexical access is suggested by the facilitatory semantic priming evident for all subjects in Experiment 3. This set

of results converges with prior studies in suggesting subject flexibility with regard to the use of assembled phonological processing. The results pose a serious challenge to single route accounts in general. Most importantly, these data speak of remarkably fined-tuned strategic adjustments in performance and suggest caution in interpreting lexical decision results without first carefully examining the specific experimental context.

REFERENCES

Andrews, S. (1982). Phonological Recoding: Is the regularity effect consistent? *Memory & Cognition, 10,* 565-575.

Battig, W., & Montague, W. (1969). Category norms for verbal items in 56 categories: A replication and extension of the University of Connecticut norms. *Journal of Experimental Psychology: Monograph,* (1969), *80.*

Besner, D., & Smith, M. (1992). Models of visual word recognition: When obscuring the stimulus yields a clearer view. *Journal of Experimental Psychology: Learning Memory and Cognition, 18,* 468-482.

Carello, C., Turvey, M. T., & Lukatela, G. (1992). Can theories of word recognition remain stubbornly nonphonological? In R. Frost, & L. Katz (Eds.), *Orthography, phonology, morphology and meaning,* (pp. 221-226). Amsterdam: Elsevier, North Holland Press.

Carr, T. H., &. Pollatsek., A. (1985). Recognizing printed words: A look at current models. in D. Besner, T. Waller, E. MacKinnon (Eds.) *Reading research: Advances in theory and practice* (Vol. 5, pp. 1-82). San Diego: Academic Press.

Coltheart, M., Davelaar, E., Jonasson, J., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and Performance VI* (pp. 535-555). Hillsdale NJ: Erlbaum. Associates

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing* (pp. 151-216). London: Academic Press.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671-684.

Davelaar, E., Coltheart, M., Besner, D., & Jonasson, J. T. (1978). Phonological recoding and lexical access. *Memory and Cognition, 6,* 391-402.

Glushko, R. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 674-691.

Hanson, V. L., & Fowler, C. A. (1987). Phonological coding in word reading: Evidence from hearing and deaf readers. *Memory and Cognition, 15,* 199-207.

Hawkins, H. L., Reicher, G., Rogers, M., & Peterson, L. (1976). Flexible coding in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 235-242.

Humphreys, G. W., & Evett, L. J. (1985). Are there independent lexical and nonlexical routes in word processing? An evaluation of the dual-route theory of reading. *The Behavioral and Brain Sciences, 8,* 68 -740.

James, C. T. (1975) The role of semantic information in lexical decisions. *Journal of Experimental Psychology: Human Perception and Performance, 1,* 130-136.

Jared, J., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in naming. *Journal of Memory and Language, 29,* 687-715.

Jared, J., & Seidenberg, M. S. (1991). Does word identification proceed from spelling to sound to meaning? *Journal of Experimental Psychology: General,* 358-394.

Johnson, N. F., & Pugh, K. R. (in press). An examination of cohort effects in visual word recognition, *Cognitive Psychology.*

Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. In R. Frost, & L. Katz (Eds.), *Orthography, phonology, morphology and meaning,* (pp. 67-84). Amsterdam: Elsevier, North Holland Press.

Kučera, H., & Francis, W. (1967). *Computational analysis of present day American English* . Providence, RI: Brown University Press.

Lukatela, G., & Turvey, M. T. (1990). Phonemic similarity effects and prelexical phonology. *Memory & Cognition, 18,* 128-152.

Lukatela, G., & Turvey, M. T. (1993). Phonological access of the lexicon: Evidence from associative priming with pseudohomophones. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 951-966.

Lukatela, G., & Turvey, M. T. (1993). Similar attentional, frequency and associative effects for pseudohomophones and words. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 166-178.

McQuade, D. V. (1981). Variable reliance on phonological information in visual word recognition. *Language and Speech, 24,* 99-109.

McQuade, D. V. (1981). Phonological recoding becomes automatic with stimulus repetition. *Language and Speech, 26,* 393-405.

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1974). Functions of graphemic and phonemic codes in visual word recognition. *Memory and Cognition, 2,* 309-321.

Monsell, S., Patterson, K., Graham, A., Hughes, C. H., & Milroy, R. (1992). Lexical and sublexical translations of spelling to sound: Strategic anticipation of lexical status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 452-467.

Paap, K. R., & Noel, R. W. (1991). Dual-route models of print to sound: Still a good horse race. *Psychological Research, 53,* 13-24.

Perfetti, C. A., & Bell, L. C. (1991). Phonemic activation during the first 40 ms of word identification: Evidence from backward masking and priming. *Journal of Memory and Language, 30,* 473-485.

Perfetti, C. A., Bell, L. C., & Delaney, S. M. (1988). Automatic (prelexical) phonetic activation in silent word reading: Evidence from backward masking. *Journal of Memory and Language, 27,* 59-70.

Pollatsek, A., Lesch, M., Morris, R. K., & Rayner, K. (1992). Phonological codes are used in integrating information across saccades in word identification and reading. *Journal Of Experimental Psychology: Human Perception and Performance, 18,* 148-162.

Posner, M., & Boies, S. (1971). Components of attention. *Psychological Review, 78,* 391-408.

Pugh, K. R., Rexer, K., & Katz, L. (submitted). Effects of orthographic neighborhood distribution in visual word recognition.

Pugh, K. R., Rexer, K., Peter, M., & Katz, L. (in press). Neighborhood effects in visual word recognition: Effects of letter delay and nonword context difficulty. *Journal Of Experimental Psychology: Learning Memory and Cognition.*

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading.* NJ: Prentice Hall.

Rosson, M. B. (1985). The interaction of pronunciation rules and lexical representations in reading aloud. *Memory and Cognition, 13,* 90-98.

Seidenberg, M. S. (1992). Beyond orthographic depth in reading: Equitable division of labor. In R. Frost, & L. Katz (Eds.), *Orthography, phonology, morphology and meaning*, (pp. 85-118). Amsterdam: Elsevier, North Holland Press.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*, 523-568.

Seidenberg, M., Waters, G., Barnes, M. A., & Tannenhaus, M. K. (1984). When does irregular spelling and pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior, 23*, 383-404.

Shulman, H. G., Hornak, R., & Sanders, E. (1978). The effect of graphemic, phonetic, and semantic relationships on access to lexical structures. *Memory and Cognition, 6*, 115-123.

Stone, G., & Van Orden, G. C. (in press). Strategic control of processing in word recognition.

Van Orden, G. C. (1987). A ROWS is ROSE: Spelling, sound, and reading. *Memory & Cognition, 15*, 181-198.

Van Orden, G. C., Johnston, J. C., & Hale, B. L. (1988). Word identification in reading proceeds from spelling to sound to meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 371-386.

Van Orden, G. C., Pennington, B. F., and Stone, G. O. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review, 97*, 488-522.

Waters, G., & Seidenberg, M. S. (1985). Spelling-sound effects in reading: Time course and decision criteria. *Memory and Cognition, 13*, 557-572.

Waters, G., Seidenberg, M. S., & Bruck, M. (1984). Children's and Adults' use of spelling-sound information in three reading tasks. *Memory and Cognition, 12*, 293-305.

FOOTNOTES

[*] *Journal of Experimental Psychology: Human Perception and Performance*, in press.

[†] Also Dartmouth College.

[‡] Also University of Connecticut, Storrs.

APPENDIX A:

Stimuli used in Experiment 2

Non-pseudohomophone condition stimuli

Experimental word stimuli:

HEEL, DOOM, CART, FLOP, SAIL, STILL, FEEL, WORD, HEAD, MOVE, DEAD, LIVE, PASS, POST, GONE, THIN, CORN, NINE, RACE, LEAST, FACE, WAKE, THESE, BEACH, SHELL, CAME, FAT, PLACE, REAL, PART, MAIN, ROAD, GAME, LAND, HEAT, DESK, FLAT, DEAF, WORM, WOOL, WARP, TOMB, HOOD, WAND, SEW, SOWN, COMB, STEAK, GROSS, FLOOD, PINT, DOLL, CROW, HOOF, COUGH, WARN, VASE, WADE, DOCK, PEST, HIKE, MATH, GREED, CHORE, GRILL, FLAG, JUNK, TILE, RUST, FLOAT, PEEL, WING, CURL, SAGE, GOAT, DISH, WASP, GLOVE, BURY, POUR, GIVE, SAYS, BREAK, TOUCH, LOSE, CHOOSE, WATCH, HEARD, BOTH, SOME, PHASE, WASH, COME, FOOT, PUT, LOVE,

Homophone fillers:

WRITE, MEET, PORE, BARE, ROLE, PAIN, HAIR, WHERE, PEAK, RAIN, MALE, PAIL, WAY, WEAK, SCENE, WHOLE

Non-Homophone fillers:

CUTE, FULL, GRAPE, GROVE, HOOK, LIKE, MASK, MONTH, SHIRT, DRESS, SHOE, RING, PEACH, SHARE, SPACE, BARN, LASH, LOAD, LIFT, MOLE

Nonwords (all non-pseudohomophonic):

BINK, BRAR, CILD, PLUB, ZATE, GRAW, FALM, FIME, PARG, PAMS, PLIN, BLAY, MOOL, RAXE, NING, FAFE, BOARB, CRECK, THEST, CLASK, COURM, KANCE, DRETS, PASH, PHANE, GLANT, GRESS, SCALB, SMICK, SROCK, TRAIZ, WALCH, BLAY, GARK, YESK, TIRT, FOID, GLAY, HAIM, TEOL, KEAR, BOUR, BAGE, JISK, SENI, PASK, VOVE, WAIG, BLOOZ, BRINP, CRILD, FEATH, DOUBS, DIGHT, FLOOG, MINTH, BEACE, GORCH, FROVE, SNILE, SPEAF, PRAGS, STELL, MEACH, MIAB, GRING, BOSK, LIPE, GARK, CALS, JEED, GOCK, YASH, MOOM, MILB, MUNT, NAIS, KIWN, RERD, DISP, SHEB, SOCH, TARL, TILK, FRUDE, DRIPE, MEAST, GREAB, GUDGE, ZINCH, JUNCH, TINSE, SCADE, SLIKE, ACOUT, SPAPE, STALM, STRUP, TOMST, TRASS, DEEE, HEIM, HERP, VISS, HOOBE, KNOZ, KNOJ, TOBE, MALP, MAPE, MASB, LOIR, SHOB, TEEP, WACE, WULD, BELGH, CRASL, KRAUD, GROST, GRELD, GRINT, BUICE, LOIST, MOURJ, WEALM, RANCE, SNORP, GOOTH, PRAIT, TRIME, WHEFE

Pseudohomophone condition stimuli

Experimental word stimuli:
Same as in Non-pseudohomophone condition.

Homophonic fillers:
Same as in Non-pseudohomophone condition.

Non-homophonic fillers:
Same as in Non-pseudohomophone condition.

Non-pseudohomophonic nonwords:

BRAR, CILD, PLUB, ZATE, GRAW, FIME, PARG, PLIN, BLAY, RAXE, NING, BOARB, CRECK, CLASK, KANCE, DRETS, PASH, GLANT, GRESS, SCALB, SMICK, SROCK, WALCH, BLAY, TIRT, GLAY, HAIM, TEOL, KEAR, BOUR, JISK, PASK, VOVE, WAIG, BRINP, CRILD, FEATH, DOUBS, DIGHT, MINTH, BEACE, FROVE, SNILE, PRAGS, STELL, MEACH, BOSK, LIPE, GARK, CALS, GOCK, YASH, MOOM, MUNT, KIWN, RERD, DISP, SHEB, SOCH, TILK, FRUDE, MEAST, ZINCH, JUNCH, TINSE, SCADE, SLIKE, SPAPE, STRUP, TOMST, HEIM, VISS, HOOBE, KNOZ, TOBE, MALP, MAPE, LOIR, TEEP, WACE, WULD, CRASL, GROST, GRELD, GRINT, BUICE, LOIST, WEALM, RANCE, SNORP, GOOTH, TRIME

Pseudohomophonic nonwords:

FONE, DOAM, HETE, FAYZE, KLEW, BOYL, TODE, COYN, BRANE, PHINE, CHUSE, FLOO, LAWD, KLAME, HOWSE, BOTE, VOYCE, SAWT, STAWL, SMYLE, RAIT, BOAL, FRUM, NIPHE, FYRE, SAIN, LUME, BRAIK, DROO, KUF, FOWND, KAVE, ROZE, GOAST, SHAWT, TOWIL, POAL, RAIK, WHEEV, PRYZE

APPENDIX B

Stimuli used in Experiment 3

Orthographically and phonologically similar pairs:

SAVE-WAVE, DONE-NONE, RUSH-GUSH, GOOD-WOOD, CARD-HARD, YARN-BARN, LIGHT-MIGHT, TON-WON, GULL-LULL, LORD-FORD, MATCH-PATCH, KID-BID, ROSE-HOSE, NEAR-REAR, HINT-TINT, MAID-RAID, SO-NO, DOVE-LOVE, PUNT-HUNT, TOUGH-ROUGH, TAR-FAR, FIVE-DIVE, HOST-POST, COW-VOW, RASH-DASH, CUT-BUT, HAND-LAND, TOMB-WOMB, FEW-PEW, BAT-HAT, DOWN-GOWN, FAST-PAST

Orthographically similar but phonologically dissimilar pairs:

HAVE-CAVE, BONE-GONE, HUSH-BUSH, FOOD-HOOD, WARD-LARD, EARN-DARN, EIGHT-FIGHT, CON-SON, DULL-PULL, WORD-CORD, CATCH-WATCH, AID-RID, NOSE-LOSE, DEAR-WEAR, MINT-PINT, PAID-SAID, GO-DO, MOVE-COVE, AUNT-RUNT, COUGH-DOUGH, BAR-WAR, HIVE-GIVE, LOST-MOST, NOW-LOW, CASH-WASH, PUT-NUT, WAND-SAND, BOMB-COMB, SEW-NEW, CAT-OAT, SOWN-TOWN, EAST-LAST

Semantically related pairs:

SHIRT-PANTS, SOCK-SHOE, TABLE-CHAIR, HAMMER-NAIL, BLACK-WHITE, CAT-DOG, LION-TIGER, APPLE-ORANGE, PEACH-PEAR, LEMON-LIME, NICKEL-DIME, SILVER-GOLD, POT-PAN, FORK-KNIFE, IRON-STEEL, HOT-COLD, RIVER-SEA, COFFEE-TEA, OAK-PINE, FLOOR-ROOF, ARM-LEG, HAND-FOOT, EYE-EAR, PISTOL-GUN, CAR-TRUCK, BAT-BALL, HORSE-WAGON, PEA-CARROT, LETTUCE-TOMATO, COTTON-WOOL, BOW-ARROW, DOVE-ROBIN

Pairs containing non-pseudohomophonic nonwords:

FOAT-BOOK, CAMB-LAMB, GIRE-GOAT, TILK-MILK, CALE-LAKE, GREE-TREE, RASK-KING, YOLE-LINT, PASH-MELT, FEST-TEST, BOUR-LIP, TILK-HELP, CIVE-LIVE, GASE-SCREW, PICE-PENNY, SHOON-SPOON, DITE-MUSK, GARE-CLOCK, FAND-FLUG, STED-LUCK, HENT-SODA, AUBE-CUBE, CHROW-THROW, KEST-TIGHT, BARO-SPIN, BOOF-SKATE, MOARD-BOARD, LIBE-ROUGH, DITE-PRINT, KINE-TIME, GEAL-WIRE, FIME-PLANT, NIRE-HIRE, CREM-FLICK, MISK-BLAST, MISH-HOLE, TOND-POND, BOCK-ROCK, PEAN-STAFF, VOMA-BASE, TOOP-HOOP, FUNE-FAULT, MALK-SWITCH, PLUST-DISK, DILM-LAMP, CODEL-MODEL, YATE-MATE, BUND-GRASS, ARCH-FOST, OIL-FOSH, PLACE-PITE, ROUTE-ZETH, LUNCH-DUNCH, SCREEN-BLID, KEY-FUT, KNOB-CLUP, STAND-LASP, THIN-TRIN, SCOOP-BLOP, TAP-BAP, TRUST-TROCK, FIX-RIX, GRAPH-TEOL, FLOAT-SLOAT, CRUSH-TOPE, LEAN-FEAN, CLAY-CHAY, HAIR-[illegible], KITE-DITE, BILL-SILE, TASTE-VASTE, GLASS-JISK, FIELD-CANK, DREAM-COSS, FAKE-TISA, TOSS-WOSS, WHEEL-NAND, MAZE-TIST, BIRD-ALKU, DEEP-MEEP, POOL-CRUNK, DECK-MECK, NET-MISEN, BUNCH-FALET, BAY-LOND, CURSE-REASY, PICK-JICK, SAFE-CROTE, BLEED-CLEED, FLAUNT-KACO, MARN-HARL, KIRM-DIRM, LURGE-FOUN, RIMP-LODY, VOCK-YOCK, PILM-DRAVE

Pairs containing pseudohomophonic nonwords:

BOOK-BOAL, GOAT-BOTE, LINT-TOAN, MELT-DEEL, HELP-DOAM, SCREW-BRANE, MUSK-DORE, LUCK-CLENE, SODA-FOAN, SPIN-GRONE, SKATE-GURL, TIME-LERN, HOLE-MEEL, BASE-NIFE, DISK-FOWND, GRASS-POAL, RUFE-ROUTE, SNOE-KNOB, TODE-TRUST, WYRE-CRUSH, MUNNY-HAIR, TITE-BILL, FYRE-FIELD, KLUB-MAZE, JALE-BAY, RAIK-MARN, LODY-RANE, DRAVE-SENE

Haskins Laboratories Status Report on Speech Research
1993, SR-114, 27-30

Horizontal and Vertical Views of Chinese Psycholinguistics*

Ignatius G. Mattingly†

During the past twenty years, there has been a very significant increase in research on Chinese psycholinguistics. Much work has been done by investigators in mainland China, Taiwan, Hong Kong, and elsewhere. Western researchers have also contributed, their interest in many instances aroused by Chinese students, who have entered Western graduate programs in linguistics, psycholinguistics, and psychology in increasing numbers. But nowhere has this development been manifested more directly than in the series of Symposia on Cognitive Aspects of the Chinese Language, of which this is the sixth.

Motivating much of this research has been a basic question: How similar are the psycholinguistic processses of Chinese speakers to those of speakers of Indo-European languages? China has a rich and ancient culture that developed completely independently from Western culture. Moreover, there are certain striking differences between Sino-Tibetan languages and Indo-European languages: Chinese has lexical tones, its syllable structure is relatively limited, its morphemes are mostly monosyllabic, and it lacks inflectional morphology. Again, the Chinese writing system appears to be very different from any present-day European writing system. Its symbols are complex patterns of strokes that stand for monosyllabic morphemes, not phonemes, and no word boundaries are indicated. There are differences in word order and vocabulary between written and spoken Chinese to which European languages offer no parallel. Finally, the Chinese writing system is central to Chinese culture, whereas the European writing systems are little more than traditional tools, having no deep cultural resonance. Given all these differences, are psycholinguistic processes for the two language families likely to be very similar?

One's expectations about the way this question will eventually be answered depend in great part on one's primary assumptions about human psychology. On one view—what Jerry Fodor (1983) has called the "horizontal" view—human cognition consists of a few basic and quite general functions: perception, memory, and motor control, for example. Given the vast range of heterogeneous input and output that human beings deal with, these functions are necessarily very versatile and very powerful. In this respect, humans differ greatly from nonhuman animals, who survive by virtue of various species-specific specializations that are highly efficient but very narrow. Someone holding the horizontal view would expect human linguistic communication to take many different forms, its variety limited only by obvious functional and anatomical constraints and encouraged by cultural and linguistic variation. A horizontalist would not be surprised to find that Chinese psycholinguistic processes, having developed under very different cultural circumstances, bore no great resemblance to those used by speakers of Indo-European languages.

Opposed to the horizontal view is the "vertical" view, which rejects "perception" and "memory" as false generalizations, and argues for psychological mechanisms, or "modules," specialized for particular domains. Of course, even the most thoroughgoing horizontalist would concede that such processes as color perception and auditory localization are precognitive specializations that can certainly be considered "modular." But a verticalist would claim beyond this that certain so-called higher-level processes, including most especially psycholinguistic processes, are also modular. In support of this view, the verticalist would point to properties that the language input system shares with input systems that are clearly modular: its limited domain, its mandatory operation, its "encapsulation" from information cognitively available to the hearer, and the limited cognitive access to the intermediate representations that it must compute, to name but a few (Fodor, 1983). On the vertical view, the language module is one more species-specific specialization, and our biological situation parallels those of other animals much more closely than the horizontalist supposes. The verticalist holds that, quite aside from functional and anatomical restrictions, psycholinguistic processes are very highly determined by particularities of the structure of the language module. He would thus expect to find only superfiial differences in these processes between Chinese and Indo-European languages.

Which of these views is more nearly correct? I think it is not too early to venture at least a tentative and partial answer. Tentative, because many questions remain unresolved, with respect to Indo-European languages as well as to Chinese. Partial, because some cognitive aspects of Chinese have received much more attention than others. But I think it is fair to say that many of the important findings for Indo-European languages have been essentially duplicated for Chinese. I will mention several of these.[1]

First, Chinese, like Indo-European languages, appears to be lateralized in the left hemisphere (Tzeng, Hung, Chen, Wu, & Hsi, 1986, and see the case-by-case review in Hoosain, 1991). Aphasia appears far more commonly in Chinese speakers with left hemisphere lesions than in those with right hemisphere lesions, and Chinese characters presented to the right visual field and hence processed first by the left hemisphere are reported more accurately than those presented to the left visual field and hence processed first in the right hemisphere. (e.g., Kershner & Jeng, 1972, and see Hoosain, 1991, Table 5.1., for other studies).

Again, Chinese readers, like English readers, take in information from print in successive ocular fixations, and the durations of the fixations are similar (Peng, Orchard, & Stern, 1983; Sun, Morita, & Stark, 1985).

"Word superiority" (Reicher, 1969) is found for Chinese as for European languages. A character is identified faster and more accurately if it is part of a two-character word than if it is part of a two-character pseudoword (Cheng, 1981; Liu, 1988; Mattingly & Xu, 1993). "Word inferiority" (Healy, 1976) is also found, again as in English: a radical that is part of a valid character is harder to detect than when it is part of a pseudocharacter, just as a letter embedded in familiar word is harder to detect than in a misspelled word (Chen, 1986).

The "Stroop effect" (Stroop, 1935) in which subjects, although instructed to report the color of the ink in which a word is printed, respond to a printed color name with that name, has been found for Chinese as well as for English (Biederman & Tsao, 1979).

In the naming task, response times depend on frequency for Chinese, as for Indo-European languages. Low-frequency characters with consistently pronounced phonetic components are responded to faster than those with inconsistent phonetic components, just as low-frequency English words that are regularly spelled are responded to faster than if irregularly spelled (Seidenberg, 1985). Visually similar but phonologically dissimilar character pairs have longer response times than control pairs in lexical decision (Hsieh, 1982, cited in Cheng & Shih, 1988), just as has been found for similarly spelled but phonologically dissimilar word pairs in English (Meyer, Schvaneveldt, & Ruddy, 1974).

Chinese characters, like words in Indo-European languages, are coded in short-term memory phonologically (Tzeng, Hung, & Wang, 1977). This is implied by the finding that phonologically similar lists are less accurately recalled. Moreover, it has shown for beginning readers of English that short-term recall ability is correlated with reading ability, suggesting that reading and recall rely on the same mechanism (Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979). Similar results have been found for Chinese (Ren & Mattingly, 1990).

It is perhaps not too much to say that whenever someone has seriously tried to find a Chinese parallel for some psycholinguistic result previously demonstrated for an Indo-European language, he has succeeded.

Differences in the results of psycholinguistic experiments between Chinese and Indo-European languages have of course been found as well, and Hoosain's (1991) argument for linguistic relativity relies heavily on these. But the differences are not very impressive. Many of them are most reasonably interpreted as showing the same basic mechanism responding appropriately to superficial variations. For example, since Chinese is often written vertically and English seldom is, it is not surprising that English readers show acuity differences between horizontal and vertical presentation, but Chinese readers do not (Freeman, 1980). Chinese readers can retain longer strings of digits in short-term memory than English readers, but this is probably because the names of the digits in Chinese are shorter (Hoosain, 1979, and see Hoosain, 1991, Table 4.2, for other studies). Readers make more fixations per line in Chinese than in English, probably because word-shape and word length information is available parafoveally in English writing, but not in Chinese writing (Peng et al., 1983; Sun et al., 1985).

Other differences found seem to be quantitative rather than categorical. While such a difference may mean something, it probably does not indicate a difference in mechanism. Thus one experiment found that "homophone" sentences are more inhibitory for English readers than for Chinese readers (Treiman, Baron, & Luk, 1981). This may mean that "getting at the meaning of Chinese is really more direct" (Hoosain, 1991, pp. 54-55), or merely that because homophony is ubiquitous in Chinese, readers have more experience in dealing with it. But it surely does not suggest a very different kind of reading process. Again, the Stroop effect is stronger for Chinese than for English (Biederman & Tsao, 1979). This may mean that the meaning of Chinese words is more manifest" (Hoosain, 1991, p. 45), or, conversely that "it is somehow unavoidable to process the *pronunciation* of the printed words" (Xu, p. 332), but it does not suggest a basic difference, as would be the case if the Stroop effect were found *only* for Chinese.

The most likely source of a basic difference in processing might be differences in the way orthography maps on to phonological structure. One way to describe the difference between Chinese and English orthographies would be to say that, unlike English, Chinese has no grapheme-to-sound conversion rules. The pronunciation of Chinese characters can be accessed only through the lexicon, whereas English words, at least those that are "regularly" spelled, can be pronounced just by using the rules, without lexical access (cf. Hoosain, 1991, pp. 36ff.). If this is the right way to view the matter, one might expect to find evidence at least of a different strategy, if not a different mechanism, in such tasks as naming. (This is a form of the "Orthographic Depth Hypothesis" proposed by Frost, Katz, and Bentin, 1987). But the evidence for such a processing difference is merely the finding that naming takes longer for Chinese than for English (Seidenberg, 1985).[2] This seems more like a difference of degree than one of kind. Perhaps a better account of the differences between the two orthographies is to say that the graphemes of English are a few score spelling patterns that specify phonemes; the graphemes of Chinese are the 900-odd phonetic radicals that specify syllables and the 200-odd semantic radicals that combine with them to form the characters.[3] Both the spelling patterns and the radicals have to be stored somehow, so both writing systems are "lexical." On the other hand, since both orthographies exhibit imperfect but still useful regularities, both can be said to have grapheme-to-sound conversion rules. Because there are so many more Chinese characters than English spelling patterns, naming a word in Chinese takes longer than naming a word in English, just as naming a word in English takes longer than naming a word in Serbo-Croatian, which has even fewer and far more regular spelling patterns than English does (Frost et al., 1987). But there is no good reason to think the underlying psycholinguistic process is very different in either comparison.[4]

It would appear, then, that the results of research on psycholinguistic processes of speakers of Chinese thus far provide substantial support for the proposition that these processes, though not yet well understood, are similar to those of speakers of Indo-European languages. This provides some corroboration for the vertical view. But there is still much to be learned. The present Symposium and its successors can be expected to provide much of the required evidence.

It may be that some Chinese investigators will regard the vertical account of psycholinguistic process as an attempt to force Chinese into a mould made by Western psycholinguists for Indo-European languages. But the vertical account of psycholinguistic mechanism is supposed to apply to all human languages. If it can really be shown not to work for Chinese in some respect, this will mean that the account needs to be revised or rejected. Nor does the vertical view in any sense demean Chinese culture. It simply asserts that all human beings have in common certain highly specialized mental structures and processes on which their cultures must ultimately depend.

REFERENCES

Biederman, I., & Tsao, Y.-C. (1979). On processing Chinese ideographs and English words: Some implications from Stroop-test results. *Cognitive Psychology*, 11, 125-132.

Chen, H.-C. (1986). Component detection in reading Chinese characters. In H. S. R. Kao & R. Hoosain (Eds.), *Linguistics, psychology, and the Chinese language* (pp. 1-10). Hong Kong: University of Hong Kong Centre of Asian Studies.

Cheng, C. M. (1981). Perception of Chinese characters. *Acta Psychologica Taiwanica*, 23. 137-153.

Cheng, C. M., & Shih, S. I. (1988). The nature of lexical access in Chinese: Evidence from experiments on visual and phonological priming in lexical judgment. In I.-M. Liu, H.-C. Chen, & M. J. Chen (Eds.), *Cognitive aspects of the Chinese language* (pp. 1-14). Hong Kong: Asian Research Service.

DeFrancis, J. (1989). *Visible speech: The diverse oneness of writing systems*. Honolulu: University of Hawaii Press.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Freeman, R. D. (1980). Visual acuity is better for letters in rows than in columns. *Nature, 286*, 62-64.

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 104-115.

Healy, A . F. (1976). Detection errors on the word *the*: Evidence for reading units larger than letters. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 235-242.

Hoosain, R. (1979). Forward and backward digit span in the languages of the bilingual. *The Journal of Genetic Psychology, 135*, 263-268.

Hoosain, R. (1991). *Psycholinguistic implications for linguistic relativity: A case study of Chinese*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Frost, R., & Katz, L.(1992). The reading process is different for different orthographies: The orthographic depth hypothesis. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 67-84). Amsterdam: North-Holland.

Kershner, J. R., & Jeng, A. G.-R. (1972). Dual functional hemispheric asymmetry in visual perception: Effects of ocular dominance and post-exposural processes. *Neuropsychologia, 10*, 437-445.

Liu, I.-M. (1988). Context effects on word/character naming: Alphabetic versus logographic languages. In I.-M. Liu, H.-C. Chen, & M. J. Chen (Eds.), *Cognitive aspects of the Chinese language* (pp. 81-92). Hong Kong: Asian Research Service.

Mattingly, I. G. (1985). Did orthographies evolve? *RASE remedial and special education, 6(6)*, 18-23.

Mattingly, I. G. (1992). Linguistic awareness and orthographic form. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 1-26). Amsterdam: North-Holland.

Mattingly, I. G., & Xu, Y. (1993). Word Superiority in Chinese. *Haskins Laboratories Status Report on Speech Research, SR-113*, 145-152

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1974). Functions of graphemic codes in visual word recognition. *Memory and Cognition, 2*, 309-321.

Peng, D. L., Orchard, L. N., & Stern, J. A. (1983). Evaluation of eye movement variables of Chinese and American readers. *Pavlovian Journal of Biological Sciences, 18*, 94-102.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology, 81*, 275-280.

Ren, N., & Mattingly, I. G. (1990). Short-term serial recall performance by good and poor readers of Chinese. *Haskins Laboratories Status Report on Speech Research SR-103/104*. New Haven CT USA: Haskins Laboratories.

Seidenberg, M. S. (1985). The time course of phonological code activation in two writing systems. *Cognition, 19*, 1-30.

Seidenberg, M. S. (1992). Beyond orthographic depth in reading: Equitable division of labor. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 85-118). Amsterdam: North-Holland.

Shankweiler, D., Liberman, I. Y., Mark, L. S., Fowler, C. A., & Fischer, F. W. (1979). The speech code and learning to read. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 531-545.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Sun, F., Morita, M., & Stark, L. W. (1985). Comparative patterns of reading eye movement in Chinese and English. *Perception and Psychophysics, 37*, 502-506.

Treiman, R. A., Baron, J., & Luk, K. (1981). Speech recoding in silent reading: A comparison of Chinese and English. *Journal of Chinese Linguistics, 9*, 116-125.

Tzeng, O. J., Hung, D. L., Chen, S., Wu, J., & Hsi, M. S. (1986). Processing Chinese logographs by Chinese brain damaged patients. In H. S. R. Kao, G. P. Galen, & R. Hoosain (Eds.), *Graphonomics: Contemporary research in handwriting* (pp. 357-374). Amsterdam: North Holland.

Tzeng, O. J., Hung, D. L., & Wang, W. S.-Y. (1977). Speech recoding in reading Chinese characters. *Journal of Experimental Psychology: Human Learning and Memory, 3*, 621-630.

Xu, Yi (1992). Review of Hoosain (1991). *Language and Speech, 35*, 325-340.

FOOTNOTES

[*] Remarks at the opening of the Sixth International Symposium on Cognitive Aspects of the Chinese Language, Taipei, Taiwan, September 2-4, 1993

[†] Also University of Connecticut, Storrs.

[1] I should at once express my debt to Rumjahn Hoosain's recent book, *Psycholinguistic implications for linguistic relativity* (1991), in which most of the relevant research is summarized, and to the review of this book by Yi Xu (1992). I should also say that Hoosain's own conclusions would probably disagree with mine, as is indeed suggested by his title.

[2] Xu (1992) questions this finding on the ground that the Chinese and the English subjects in Seidenberg (1985) may not have been at comparable educational levels.

[3] On Chinese writing as a syllabary system, see Mattingly (1985, 1992) and DeFrancis (1989).

[4] For further discussion of the Orthographic Depth Hypothesis, see Frost and Katz (1992) and Seidenberg (1992).

Haskins Laboratories Status Report on Speech Research
1993, SR-114, 31-80

Learning to Perceive the Sound Pattern of English*

Catherine T. Best†

Language lies at the heart of human cognitive and social development. Infants, who are by definition "without language," become speaker-hearers of particular languages within their first few years, through their experience with the speech of their caregivers and other significant people in their environment. The foundation for the emergence of language proper is the infant's discovery of sound-meaning correspondences in the utterances produced by those significant people. Social and physical context provide support for the semantic meaning of an utterance, although determining the specific referent of an unknown word from non-linguistic context alone may be no simple task (see Quine, 1960). The present discussion, however, will focus on the other side of the sound-meaning relation, the sound pattern itself. It still far from clear how the infant comes to recognize in the stream of connected speech the sequence of consonants and vowels that may underlie the diverse pronunciations of a given word in different sentences, by different speakers, and under different speaking conditions (e.g., in rapid casual speech versus slow, exaggerated infant-directed speech). Presumably, these accomplishments are built on the infant's prior abilities to discriminate and classify the audible properties that correspond to various levels of organization in speech, e.g., consonants and vowels (phonetic segments), rhythmic stress patterns, prosodic phrases, and so forth.

It is these perceptual abilities for handling the "surface phonetic structure" of speech that are the primary concern of this chapter. In particular, we will focus on how the infant's experience with a particular language begins to influence perception of consonant and vowel contrasts that fall outside the phonetic inventoryemployed by that language. Developmental changes in perception of such non-native contrasts can provide important insights about the aspects of the native phonological system to which infants are becoming attuned as they gain experience with native speech. The central goal of this chapter is to describe and provide evidence for a model of how language-specific experience influences infants' and adults' perception of non-native phonetic contrasts. The model is the Perceptual Assimilation Model of cross-language speech perception.

First, however, we must briefly review the basic pattern of developmental change in perception of non-native phonetic contrasts, and describe the phonetic and phonological organization in spoken language that the infant must come to perceive. Following that introduction to speech and its perceptual requirements, we will consider two major theoretical perspectives that might be extended to account for language-specific developmental changes, to provide a backdrop for the presentation of the Perceptual Assimilation Model.

Infants' perception of phonetic properties in speech

Young infants can discriminate a wide range of phonetic contrasts between consonants (e.g., [b] vs. [d]) or between vowels (e.g., the vowels in *boot*

Preparation of this chapter was supported by an NIH Research Career Development award from the National Institute on Deafness and Communication Disorders (DC00045). Research from my lab that is described herein was also supported by NIH research grant DC00403 to the author and by program project grant HD01994 to Haskins Laboratories. Many thanks are due to Alice Faber, Andrea Levitt, Steven Braddon, Janet Werker, Peter Jusczyk and Carolyn Rovee-Collier for their helpful comments on an earlier draft of the chapter, to Alice also for managing the infant lab in my near-absence, and to a wonderful group of research assistants who kept the lab running smoothly while I wrote: Janet Calderòn, Sandy Chiang, Ron Dewitt, Domenica Giancolo, and Alyssa Wulf. Thanks also to Steve for doing more than his fair share of child care, and to my daughters Aurora and Vanessa for being patient.

vs. *book*), whether or not the tested phonetic features are employed linguistically by the ambient language. But by adulthood, in fact by much earlier in development, experience with the native language comes to exert some rather striking effects on the perception of phonetic contrasts. The experiential influence is particularly apparent for perception of contrasts that are not part of the native language's phonological system. As will be explained more fully in the next section, the phonological system refers to the rules by which a given language employs certain phonetic differences as linguistic contrasts that can convey differences in word meanings. It treats certain other phonetic differences as linguistically equivalent, and yet other phonetic features as non-permissible altogether even though the same features may be used linguistically by some other language. Mature listeners often have substantial difficulty discriminating and categorizing phonetic contrasts which are not part of their own phonological system, but young infants from the same language environment have no difficulty discriminating those same contrasts. Effects of language-specific experience emerge in speech perception during the second half of the infant's first year, and are clearly evident by 10-12 months for perception of many non-native consonant contrasts (see reviews by Best, 1984, 1993, in press, a; Werker, 1989, 1991; Werker & Pegg, 1992).

Why and how does experience with the native language come to shape the perception of the phonetic properties of speech in this manner? How do infants become familiar with the sound system of their native language, and how does that process subsequently shape perception of unfamiliar consonants and vowels from languages not heard before? Infants' initial experience with their language begins with only the surface phonetic patterns of spoken utterances, but ultimately they must use that input to develop knowledge of the underlying semantic concepts and syntactic rules of the language. Thus, the first inroads the infant makes into discovering the systematic structure of the language take place at some level of its sound system. Many believe that this discovery process commences at the prosodic level.

Recent research on prosodic bootstrapping—the notion that conversational speech (particularly infant-directed speech) provides converging intonational and rhythmic markers that guide infants' attention to clause and phrase boundaries in speech—has made important advances in our understanding of how infants may discover the boundaries of syntactic units at varying levels (e.g., Gleitman, Gleitman, Landau, & Wanner, 1988; Hirsh-Pasek, Kemler Nelson, Jusczyk, Wright Cassidy, Druss, & Kennedy, 1987; Jusczyk & Kemler Nelson, in press; Kemler Nelson, Hirsh-Pasek, Jusczyk, & Wright-Cassidy, 1989; Morgan, 1990). However, prosodic bootstrapping may not help the infant so much with segmenting sound at the word level. Broad prosodic markers do not consistently specify word boundaries in continuous speech (cf. Gerken & Mcintosh, 1993; Jusczyk, Cutler, & Redanz, 1993), especially in languages like French which lack syllabic stress alternation patterns like those found in English. But word boundaries are often marked by characteristic differences in the exact way that the surrounding consonants and/or vowels are pronounced (e.g., aspirated [t] and reduced "uh" vowel in *citrus*but not in *sit Russ*), phonetic characteristics to which even very young infants appear to be sensitive (Christophe, Dupoux, Bertoncini, & Mehler, submitted; Hohne & Jusczyk, 1992). Thus, word-segmentation may be aided not so much (or not only) by prosodic bootstrapping but more by what might be called phonetic bootstrapping.

It is the infant's attention to this sort of detailed phonetic information that would seem to be most relevant to the discussion of how language-specific experience begins to influence perception of consonants and vowels, also referred to as phonetic segments. A basic premise of this chapter is that infants make use of surface phonetic details to discover the more abstract phonological properties of their native language. As will be described more fully in a subsequent section, the phonological system refers to the inventory of phonetic segments that a given language employs to convey meaningful differences among words. This inventory is organized systematically and hierarchically around multiple contrasting phonetic features that define linguistically important relations among phonetic segments. The systematicity of a language's phonological system makes possible the vast expansion of vocabulary that takes place in early childhood, and somewhat later serves as the linguistic framework for the child's acquisition of reading and writing abilities. But the relation between the surface phonetic details of utterances and the more abstract phonological system of a language is not always transparent, in part because of contextually-determined differences in the phonetic details of consonants and vowels, and other effects such as speaker and speaking rate differences in pronunciations. Thus, in order to learn the sound pattern of the ambient language sufficiently to determine sound-meaning relations, the infant must begin to untangle the complex relationship between the surface phonetics and the underlying phonological system, at least to some approximation.

To provide a foundation for considering developmental changes in speech perception, we will turn now to an overview of the hierarchical nesting of linguistic information conveyed in the speech signal. We will focus in particular on the relationship between the lower-order patterning at the surface phonetic level of speech and the more abstract, higher-order organization at the phonological level of a given language. Differences in the sound patterns of different languages reflect differences not only in their inventories of consonants and vowels, but also especially in the patterns by which they relate phonetic details to phonological structure. It is the relationship between phonetic details and phonological organization that is most germane to understanding the effects of language experience on the perception of non-native speech sound contrasts. Any theory of the acquisition of native language sound patterns, and of the perception of those patterns, must be able to take into account the sound structure of the spoken message and the observations of language- and dialect-specific differences in that structure.

The structure of the spoken message

When we convey a spoken message to a listener, the utterance we produce via the audible, and to some extent visible, articulatory movements of our vocal tract is organized according to the multiple levels of linguistic structure of the language we speak (the property of dual structure: Hockett, 1963). That is, the spoken utterance concurrently reflects the organizing of sound into words, the syntactic organization of those words into the larger units of noun, verb, or other phrases, and the superordinate syntactic organization of phrases into clauses, one or more of which may comprise a sentence. At the same time, prosodic organization is evident in the intonation, temporal patterns, and amplitude changes that provide a common carrier for the words at the phrase, clause and sentence levels, and serve to signal linguistic stress, pragmatic emphasis and emotional tone. But there is also nested structure if we look in the opposite direction, below the level of individual words. A word is composed of one or more units of meaning, referred to as morphemes, e.g., the word *incomplete* contains the stem morpheme *complete* plus the negation prefix *in-*. Morphemes are comprised of one or more syllables, each made up of consonants and vowels, which are defined in standard linguistic analysis as phonological segments.

Phonological patterning

Phonological segments are the smallest units of the language-specific grammatical system. They are themselves composed of phonetic features, the matrix of articulatory/acoustic properties that characterize the way a given phoneme is produced. These properties are described according to a universal set of distinctive feature contrasts by which one segment can differ critically from all others (e.g., Jakobson, Fant, & Halle, 1963; for an introduction to phonetics, see Catford, 1988; Ladefoged, 1982). For example, the consonants and vowels in the word *incomplete* may be broadly transcribed to correspond to phonemic segments as /ɪnkamplit/. However, additional phonetic details that are present in the actual production of the word can be represented in a narrow phonetic transcription as [ɪŋkəmpʰl̥it̚]. The narrow transcription indicates that the /n/ preceding the /k/ is actually produced as a nasalized constriction [ŋ] near the soft palate at the back of the mouth, rather at the alveolar ridge behind the upper front teeth [n]. The vowel in the second, unstressed syllable is the reduced vowel schwa [ə], which is somewhat like the "uh" ([ʌ]) in *butter*, but shorter in duration. The /p/ is produced with breathy aspiration [pʰ], which causes the following /l/ to be devoiced [l̥]. And the tongue-tip closure for the final /t/ is not audibly released at the end of the word [t̚]. (For an introduction to phonology see Kenstowicz & Kisseberth, 1979).

The phonology of a language is the set of systematic constraints the language places on the sound patterning of its consonants and vowels. To begin with, every language employs but a subset of all humanly-producible consonant and vowel sounds to produce minimal phonological contrasts in word meanings. As an illustration of minimal contrast, English uses /b/ and /p/ to differentiate the meaning of words that are matched in their other phonemic elements, such as *bat* vs. *pat* . Likewise, the vowel contrast /ɪ/-/ɛ/ distinguishes the minimally contrasting words *pit-pet* (/pɪt/-/pɛt/). However, modern English lacks the throaty fricative at the beginning of the Yiddish word *chutzpah*.

The phonology of a language also includes contextually-determined allophonic variations in the phonetic details of a given phoneme produced in different surrounding contexts. For example, in English the /p/ in *pan* is produced with aspiration and a long lag before voicing starts after the release of the bilabial closure, denoted phonetically as [pʰ]. But the /p/ in *span* is produced with a much shorter voicing lag and without aspiration, denoted as the allophone [p]. However, this difference in pronunciation does not signal a phonological contrast in English. Phonological analyses of the range and constraints on allophonic variants reveal which one is the underlying phonological form, and which others are the variants of that underlying form. In this case, [p] is a variant of underlying [pʰ]. There are no English minimal word pairs whose meaning is differentiated phonologically solely by the /p/-/pʰ/ difference.

Certain other contextually-determined effects on the phonetic details of segments in a spoken message result from more global changes, such as different speech rates and styles. To illustrate, the phrase *did you eat*...in slow, careful speech is typically produced with two clear /d/'s and the "ih" vowel in *did*, clear "y" and long "oo" sounds for *you*, and a clear "ee" and /t/ in *eat*.. But in rapid, casual speech the phrase may become *d'y'eat*...where the initial /d/ and vowel in *did* have been omitted, the final /d/ seems to combine with the "y" of *you* to form a "j" sound, and the long "oo" has become an unstressed schwa [ə] (e.g., Oshika, Zue, Weeks, Neu, & Aurbach, 1975; Browman & Goldstein, 1990a).

Languages also have phonotactic constraints on the distributional patterns of consonants and vowels, including permissible sequences in syllables and permissible positions that particular sounds can occupy within a syllable or word. For example, /spa/ and /mop/ (*mope*) are permissible English syllables but */psa/ and */mpo/ are not. Also, English words may end but may not begin with the velar nasal /ŋ/ (as in *song*) or may have an internal voiced palatal fricative "zh" (as in *measure*) but may not begin with this sound.

Thus, the phonological system of a language refers to the underlying linguistically-defined relations among the consonant and vowel sounds it employs. The language's use of consonant or vowel differences for contrastive differentiation of word meanings, the allophonic patterning of those phonemes, and their phonotactic distributional constraints all reflect abstract invariant properties that underlie the surface phonetic details of spoken utterances. As should be clear from these examples, the relation between the phonetic details and the phonological organization of a language is often far from a simple, transparent mapping.

To address how infants might learn aspects of the language-specific phonology from ambient speech, and how that might influence their perception of non-native phonetic contrasts, we need to briefly review next how languages differ in the ways they relate the phonetic details of speech to phonological structure.

Language differences in phonology and phonetics

An obvious way in which the sound patterns of languages differ is in their inventories of phonological segments and minimal contrasts. Although certain basic segment types seem to be universal, or nearly so, across the inventories of the world's languages, other sounds and contrasts are present only in some languages and are absent in others. Among the universally-shared phonological segments are the stop consonants /p/ and /t/ and the vowels "ah" as in *father*, "ee" as in *see*, and "oo" as in *boot*.[1] Language differences in phonological inventories are numerous, however. For example, the /l/-/r/ contrast found in the inventory of English is absent from many Asian languages, such as Japanese and Korean, as well as from a number of other languages; indeed, the English /r/ is quite rare across languages. Similarly, the English vowels in *hook* and *hawk* respectively, are lacking in Spanish, Native Hawaiian and many other languages. Conversely, English lacks the click consonants of Zulu and other southern African languages, as well as the dental versus retroflex stop consonant contrast /d̪/-/ɖ/ of Hindi (our /d/ has a tongue-tip position in-between the Hindi sounds). English also lacks the front rounded vowels /y/-/ø/ found in French, German, Swedish and elsewhere.

The neat and straightforward description of language differences in phonological inventories is seemingly complicated, however, by the fact that languages also either require or permit certain context-conditioned or free allophonic variants for at least some of their phonemes. For example, the French /r/ is characterized as a voiced uvular trill at the back of the throat, yet context-conditioning causes its surface phonetic form to become a voiceless uvular fricative when it follows a voiceless consonant, e.g., as in *quatre*, the French word for "four." Permissible differences among speakers also result in other freely varying allophones.

Allophonic variations may even, at times, appear to obfuscate claims that one language lacks a particular phoneme or contrast found in another. To illustrate, neither the dental nor the retroflex stop that contrast in Hindi are found in the English phonological inventory. Our /d/ is underlyingly a voiced alveolar stop [d]. However, a dental stop does occur *phonetically* in English speech, as an allophone of /d/ that is context-conditioned due to coarticulation (overlapping production) with adjacent dental sounds. The dental allophone occurs when /d/ is adjacent to a dental fricative e.g., in *birthday*. These observations might seem to belie the claim that only Hindi, and not English, has a dental stop in its phonological inventory. The important point, though, is that this dental form does not contrast with /d/ in English. It is a context-conditioned allophone of /d/ and is heard as /d/. The adjacent dental segment is perceived as the source of the variant property (see also Fowler & Smith, 1986; Kent, Carney, & Severeid, 1974; Krakow, Beddor, Goldstein, & Fowler, 1988; Mann, 1980, 1986; Whalen, 1983), apparently even by young infants (Fowler, Best, & McRoberts, 1990).

The discussion about language differences in allophonic patterning prompts consideration of a similar phenomenon in which different languages, and different dialects of a single language, can differ in their phonetic realizations of the "same" phonological segment. If the phonetic details differ, then on what basis is the underlying segment in such cases "the same," in at least some crucial way? This question is more problematic for the cross-language case, but several observations suggest that underlying identity of segments, or at least close similarity, may often be a reasonable assumption nonetheless (see also Flege, 1987, in press). For one thing, the phonetic feature matrix that defines a given phonological segment includes only those features critical for distinguishing it from other segments in a language's phonology. Allophones are encompassed in the definition because they vary on non-critical features. Thus, English and Spanish both have the phonological segment /p/ even though it is often aspirated in English but never in Spanish. It is important to note, however, that listeners are quite sensitive to foreign accent in their native language, suggesting that listeners may nonetheless detect such sub-phonemic differences. Findings indicate that while some of the sensitivity to foreign accent is attributable to prosodic differences, for at least some cross-language segmental similarities the phonetic differences between the corresponding native and non-native segments are also perceptible (e.g., Flege, 1984, in press; Flege & Eefting, 1987; Flege & Fletcher, 1992).

Cross-language identity and similarity are corroborated by the phonological forms speakers use when learning a new language with unfamiliar pronunciations, as when a Spanish speaker's initial pronunciation of English *pit* may sound like *beet* because he or she uses the Spanish unaspirated /p/ and an "ee" vowel because Spanish has no "ih" sound. Cross-language segmental similarities are also suggested by the phonological forms speakers of one language give to loan-words from another language (see also Silverman, 1992). For example, the French *calorique*, pronounced with an unaspirated /k/, an uvular trilled /r/ and the vowels "ah," "o" and "ee," has been adopted into English as *caloric* and pronounced with an English aspirated /k/, English /r/, and unstressed schwa [ə] in the first and final vowel positions.[2] Moreover, similar sorts of phonological substitutions are seen in pidgins and creoles, inter-languages which result from social contact between two independent language groups, and which often derive only from spoken forms at least in their early stages (e.g., Holm, 1988; Romaine, 1988). Finally, the patterns by which listeners label non-native segments, not surprisingly, provide further converging evidence about cross-language segmental similarities, as will be described later.

By comparison to the cross-language case, the segmental identity issue seems relatively straightforward for the cross-dialect case, at first glance. For mutually intelligible dialects, the vocabulary, the grammar (phonology, morphology, syntax) and even the written forms are typically nearly identical between dialects. In this case, there is no doubt about phonological identity between corresponding segments in the dialects, even though they differ in some phonetic details. Here again, listeners nonetheless detect dialectal accent easily, and show differential sensitivity to phonetic differences among segments in the native vs. non-native dialects (see Faber, Best, & DiPaolo, 1993).

Numerous examples of cross-dialect phonetic variants of underlying segments can be found in languages. On portions of Long Island in New York, words such as *long* are pronounced with a final /g/, although the final /g/ is omitted elsewhere in the U. S. To take an example from another language, the nasalization of vowels in Canadian French commences later into the vowel than in continental French (van Reenen, 1982). Paralleling another between-language difference, one dialect may lack a phonological contrast found in other dialects of the same language (or found historically in the language), a situation termed a "merger" of the contrast. For example, English speakers from Canada, western U.S., and areas of the midwest U.S. fail to produce or reliably label the "aw"-"ah" difference, as in *hawk-hock*, a vowel difference that is maintained in the northeast U. S. (e.g., Di Paolo, 1992). Similarly, Texans have merged the "ih"-"eh" difference before /n/, pronouncing *pin* and *pen* as homonyms (both like *pin*).

Sometimes, a merger is not absolute, but rather is a "near-merger" (see Faber, Di Paolo, & Best, submitted; Labov, 1974; Labov, Karen, & Miller, 1991). In a near-merger, a phonological contrast found elsewhere in the language is no longer evident in a given dialect, but productions of the near-merged sounds still show reliable acoustic differences and/or the contrast reappears in a subsequent sound change in the dialect. One such historical reversal occurred in early Modern English. The vowels in words like *meat-mate*, which had merged earlier, later re-established different pronuciations when the *meat* class but not the *mate* class vowels merged with the vowel in words like *meet* (the *meat-meet* merger still stands today) (Labov, 1974). As an example of near-merger in current American English, /r/ is dropped after "ah" in some Boston dialects. Thus, word pairs such as *cod-card* are produced as near-homonyms (Costa & Mattingly, 1981). A similar effect is found in many dialects of British English. A near-opposite pattern occurs in Brooklyn, where speakers add /r/-color to the "aw" sound, pronouncing *sauce* like *source* (Labov, Yaeger, & Steiner, 1972). In Albuquerque and the Salt Lake Valley, vowel pairs such as "ee"-"ih" and long "oo"-short "oo" (as in *boot-book*) show near-merger in the context of a following /l/. That is, word pairs such as *pool-pull* and *heel-hill* are pronounced as near-homophones (Di Paolo & Faber, 1991; Faber, 1992; Labov, Yaeger, & Steiner, 1972).

To return to cross-language differences, languages often differ in the phonotactic constraints they place on the sequences and word positions permitted among the segments in their inventories. As an illustration, English does not permit the "zh" sound word-initially, but a number of other languages do, as in the French word for magazine, *journal*, and the Russian word for woman, *zhenshchina*. Likewise, English disallows "ng" ([ŋ]) in word-initial position, but that position is allowed in Vietnamese, as in the name *Nguyen*. On the other hand, stop consonants such as /p/, /t/, /k/ can occur in initial but not in final positions in Mandarin Chinese words and syllables; in English they can occur in either position. Finally, English phonotactics disallows certain phoneme sequences in syllables that are nonetheless permissible in other languages, such as */psa/ (e.g., in Greek), */mpo/ (e.g., Chaga), and */dzva/ (Polish).

In addition, the types of phonological alternations present in one language may be absent from others. As an example, Turkish uses a phonological principle of vowel rounding harmony within words, whereby the vowels in a word must agree in whether they have lip-rounding (e.g., "o" and long "oo") or not (e.g., "ee" or "ih"). Thus, the possessive form of *dere*, the word for river, is *deresi̲* but the possessive form of *boru*, the word for pipe, is *borusu̲*. English, of course, does not require any sort of vowel harmony. Other languages have a rule of vowel epenthesis to maintain a regular pattern of consonant-vowel alternation, whereby a vowel is inserted between any adjacent consonants. For example, pluralizing the Chuckchee word for river *wejem* by adding the plural morpheme *-ti* results in *wejemet* and not **wejemti* because the /m/ and /t/ must be separated by a vowel (the final *i* is deleted through a separate phonological rule). As a final example, some dialects of Spanish have a rule of spirantization by which voiced stop consonants /b/, /d/, /g/ become voiced fricatives following a vowel, as in the pronunciation of *nada*, the word for "no," with a dental fricative instead of a /d/. It is interesting to note that the early words of young English-learning children often display phonological constraints that are absent from adult English, but similar to rules found in other languages. For example, complete vowel harmony is evident in "baba" for *bottle* and "dada" for *daddy*, while vowel epenthesis is evident in "buhlue" for *blue*. However, children's early phonologies sometimes also display other constraints that are seldom if ever seen in adult phonologies, such as the childish consonant harmony constraint by which *doggy* is produced as "dawddy" or *ducky* as "gucky."

Language differences in phonological inventories and in the phonetic properties of identical or similar phonological segments are the primary aspects of phonology with which we will deal in the remainder of the chapter. These are the aspects of speech most likely to be relevant to considering the lowest-level invariants of native language structure that infants may initially recognize in the consonants and vowels of the ambient language. But how is it that the infant moves from the surface phonetics to the underlying phonology? And how might the infant's progress on this front be reflected in changing perceptual responses to non-native phonetic patterns?

On accounting for developmental changes in perception of phonetic information

Two comprehensive, but radically different, theoretical approaches stand out in the scientific literature as providing possible accounts of how infants become attuned to the phonetic properties of their native language and begin to sort out the phonetics-phonology relations. The first approach is Noam Chomsky's linguistic theory of the grammatical structure of language and of its implications for language acquisition. Chomsky's premise of an innate Language Acquisition Device (LAD) is probably the most well-known and widely-accepted nativist perspective on language development. It is probably less widely known that his LAD was meant to apply to phonological as well as syntactic processes. The second is a psychological theory that is rarely applied to language or its development, James and Eleanor Gibson's ecological perspective on perception. Their notion of perception as information pickup would suggest, as an alternative to an innate linguistic device, that perceptual learning may be the means by which language experience affects perception of native versus non-native phonetic information.

To provide the foundation and rationale for the Perceptual Assimilation Model of language-specific effects on speech perception to be presented in the subsequent section, this section of the chapter will critically examine Chomsky's and Gibson's theoretical approaches. It will be argued that while Chomsky's theory has provided important insights about the grammatical structure of language, including its phonological properties, some of his basic claims about the phonetics-phonology relation have not been supported by subsequent work in phonology. More important, difficulties with his nativist perspective on development lead me to reject that view as an approach to understanding the development of language-specific effects on speech perception, in favor of the perceptual learning approach outlined by the Gibsons

Following this theoretical discussion, PAM will be developed as a perceptual learning account of listeners' perception of non-native contrasts according to their phonetic similarities and dissimilarities *vis a vis* native phonological categories. The model is based on the principles of information pickup and perceptual learning put forth in the ecological theory of perception, as applied to listeners' recognition of language-specific relations between surface phonetic details and the underlying phonological principles that have been characterized by linguistic research. The model will be discussed in light of recent cross-language perceptual findings with infants and adults, from my own and others' laboratories. In addition, PAM's implications for the development of phonological knowledge about the native language will be considered.

Let us turn now to our evaluation of Chomsky's proposal about language acquisition, and of the Gibsons' theory of perception and perceptual learning. This discussion provides the groundwork for PAM.

Chomsky and the Language Acquisition Device

To set the stage, consider a quote from Chomsky's *Language and Mind* (1972), which illustrates his reasoning about the need for a language acquisition device. This particular passage was chosen because of its emphasis on the role of the LAD in phonological development.

> "[W]e can provide an explanation for a certain aspect of perception and articulation in terms of a very general abstract principle, namely the principle of cyclic application of rules. It is difficult to imagine how the language learner might derive this principle by 'induction' from the data presented to him. In fact, many of the effects of this principle relate to perception and have little or no analogue in the physical signal itself, under normal conditions of language use, so that the phenomena on which the induction would have been based cannot be part of the experience of one who is not already making use of the principle.... Therefore, the conclusion seems warranted that the principle of cyclic application of phonological rules is an innate organizing principle of universal grammar that is used in determining the character of linguistic experience and in constructing a grammar that constitutes the acquired knowledge of language." (Chomsky, 1972; p. 45)

As indicated, a core premise of Chomsky's theory is that humans possess an innate biological specialization for learning language. This specialization is devoted solely to determining the specific grammatical structure of the native language, within the innately-specified constraints on possible human grammars, on the basis of spoken input. The biological device, the LAD, is endowed with the universal grammar, that complement of grammatical functions found universally across languages. Thus, it includes the mechanisms that generate the language-specific rules by which the surface phonetic representations of utterances are derived from the underlying deep structure, or abstract phrasal organization of intended meaning. Cross-language similarities in the structure of children's early grammatical constructions, their common phonological simplifications in pronouncing early words, and the disparity between those childish constructions and the grammars of the adult languages, are taken as evidence for an innate biological specialization for language acquisition. The LAD makes possible the child's construction of a representation of the grammatical system of the native language, which includes the phonological rules by which sound and meaning are related, as can be seen in the following quote.

> "[T]he child constructs a grammar—that is, a theory of the language of which the well-formed sentences of the primary linguistic data constitute a small sample.... A child who is capable of learning language must have (i) a technique for representing input signals, (ii) a way of representing structural information about these signals, (iii) some initial delimitation of a class of possible hypotheses about language structure, (iv) a method for determining what each such hypothesis implies with respect to each sentence, (v) a method for selecting one of the (presumably, infinitely many) hypotheses that are allowed by (iii) and are compatible with the given primary linguistic data." (Chomsky, 1965, p. 25-30)

Although his work on syntax is more extensive and more widely known outside of linguistics than his work on phonology, it is important to note that Chomsky considered the phonological patterning of a language to be a component of its grammar. Therefore, the endowment of the LAD also had to include the universal set of phonetic features—the full range of possible speech sound features from which all languages select a subset for the surface phonetic representation of utterances. The next quote, from *The Sound Pattern of English* (Chomsky & Halle, 1968—henceforth referred to as SPE), describes the predicted effects that knowledge of a particular language should have on the perception of phonetic features in speech.

> "The hearer makes use of certain cues and certain expectations to determine the syntactic structure and semantic content of an utterance.... A person who knows the language should 'hear' the predicted phonetic shapes.... Notice, however, that there is nothing to suggest that these phonetic representations also describe a physical or acoustic reality in any detail.... Accordingly, there seems no reason to suppose that [even] a well-trained phonetician could detect such contours with any reliability or precision in a language that he does not know..." (Chomsky & Halle, 1968, p. 24-25)

Thus, Chomsky and Halle posit that a listener's perception of phonetic patterns is determined by the phonological component of the specific grammar of his or her native language, *once the listener knows the language*. But if only a person who knows the language hears the phonetic shapes predicted by the grammar—the meaningful contrasts and phonetic equivalencies within its phonological component—then how should those same phonetic patterns perceived by someone who does not know the language? More specifically, How does perception of the phonetic details of an unknown language differ between a listener who knows at least one language (i.e., knows a different language-specific grammar) and a listener who has not yet learned a first language (i.e., does not yet know a particular grammar)? How, indeed, does the first language learner acquire the native phonology, based on the spoken input from his or her language environment?

The answer to the last question, according to Chomsky, is that the LAD helps young children to determine the language-specific grammatical operations that relate the surface phonetic forms of native utterances to their underlying phonological, syntactic and semantic representations. Because young infants innately possess the set of universal phonetic features, they should perceive the full range of possible surface phonetic contrasts in non-native as well as in native speech. In this way, they remain open to learning whichever language is presented to them. But why, then, don't adults and older children also perceive the universal phonetic features in non-native speech? The brief treatment of this issue in SPE points to the answer. It cannot be that mature language users have somehow lost the universal phonetic features with which they were born. Rather, it must be that for them the language-specific grammatical rules they have come to possess necessarily translate the surface phonetic features of utterances to the underlying phonological representations that are in accord with the grammatical principles of their language(s). That is, once the child has determined the rules of the language-specific grammar, s/he will "hear" the phonetic shapes predicted by the phonological component of that grammar.

This process would not constrain young infants' perceptions because they have not yet accrued sufficient language input to determine the underlying language-specific phonological representations of the ambient language's grammar. The LAD and its universal grammar are, nonetheless, present and operating even in the young infant. Its function in phonological development at this early stage is to construct the underlying grammar of the phonological component of the language by generating and testing hypotheses that could account for the observed patterning of the surface phonetic details in ambient speech.

To understand how this was expected to take place, we must briefly examine Chomsky and Halle's basic assumptions about how phonetic details relate to phonological representations. The classic view of SPE was that each consonant and vowel in an utterance is a discrete segment, represented phonologically as a feature matrix of all and only those phonetic features that distinguish it from all other segments in the language's inventory. The role of the phonological component of the grammar is to assign a language-appropriate phonetic feature matrix for the surface structure of each utterance generated by the syntactic component of the grammar. Thus, the phonological mapping to phonetic features is a part of the language-specific grammar. But the phonetic features are assumed to be binary, abstract, and timeless representations, even though their physical articulatory instantiations extend over time and space and show graded variability. That is, each static phonetic feature in a segmental matrix has only a positive (+) or a negative notation (-); the values for all features hold absolutely and concurrently in a segmental representation that has no time dimension. These static, binary feature specifications of the surface phonetic representation are automatically translated into the continuous, scalar articulatory details of real utterances, with temporal and spatial extent, by the universal grammar. That is, the translation to physical articulations is *not* part of the language-specific grammar. For these reasons, phonological representations do not incorporate all of the actual articulatory details associated with particular physical instantiations, such as the full range of details for specific dialectal or allophonic variants of a given segment. The latter sorts of detailed descriptions might be provided (by phoneticians) to fully characterize allophone-specific, dialect-specific, or even language-specific properties of utterances. But these would not be part of the language-specific grammar, and so are not essential descriptions of phonological segments, which are abstract. Phonological segments represent the functional patterning of sound by the language's grammar, and therefore are blind to allophonic or dialectal differences, which are phonologically equivalent in the underlying representation.

It is important to point out, however, that this segmental or linear view of phonology as propounded in SPE has largely been supplanted more recently by nonlinear or autosegmental phonology (e.g., Archangeli, 1988; Archangeli & Pulleyblank, in press; Clements, 1985; Keating, 1988, 1990; McCarthy, 1988; Prince & Smolensky, 1993; Sagey, 1986; for an introduction to autosegmental phonology, see Goldsmith, 1976). The nonlinear approach has developed in response to several difficulties with the classic linear model's handling of certain aspects of phonological patterns and phonetic implementations across languages. For one, the SPE claim that all features are binary fails to account for certain phonological processes; the nonlinear approach instead recognizes multivalent settings for certain phonological features. For another, the exclusively segmental domain of the SPE model failed to coherently incorporate certain effects of stress patterns, intonation, and syllable structure (phonotactics) on segmental properties. These effects are handled in nonlinear accounts by assuming instead that segments, stress, tonality, and syllable organization are distinct but interacting subcomponents of the phonology (e.g., Ito, 1986; Leben, 1978; McCarthy, 1986, 1989; Pierrehumbert & Beckman, 1988).

Another common phonological pattern is that phonetic features of one segment often "carry over" to other segments in an utterance, e.g., vowel harmony, context-conditioned allophones. Because SPE assumed phonetic features are linked to individual phonological segments, these phenomena required a proliferation of rules for moving phonetic features between segments. In nonlinear phonology, the effects follow automatically from an assumption that all features are independent of specific segments, with possible associations to one or more segmental "slots" (e.g., Cohn, 1990; Goldsmith, 1976; Inkelas & Leben, 1990; Kahn, 1980).

Finally, language- and dialect-specific differences in productions of segments with identical phonetic feature specifications call into question the SPE argument that articulatory implementation of phonological representations is automatic and universal, suggesting instead that articulatory details are part of language-specific grammar (see Fourakis & Port, 1986; Keating, 1988, 1990a, b; Mohanan, 1986). For example, the ejective stop /p'/ is released later and hence more forcefully in Navajo than in Quechua (Lindau, 1984); and nasal vowels have more delayed nasalization in Canadian French relative to continental French (van Reenen, 1982).

Although it has gone beyond the SPE model in handling certain phonetic and phonological patterns, however, the nonlinear approach has apparently retained the other basic theoretical premises of SPE. The nonlinear approach still assumes that phonological features are abstract and timeless. Moreover, nonlinear phonology proponents have had very little to say about ontogenetic development, certainly nothing that differs substantively from Chomsky's nativist assumptions (e.g., Archangeli & Pulleyblank, in press). That is, the nonlinear approaches retain, either tacitly or explicitly, the notion of an innate language acquisition device containing a universal grammar, with universal phonetic features.

However, those unquestioned assumptions, particularly certain assumptions underlying the posited innate linguistic device, raise some vexing problems. In-depth critiques of Chomsky's general theoretical framework have been offered from a linguistic perspective by Derwing (1973) and Sampson (1980), and from a psychological perspective by Bohannon, MacWhinney, and Snow (1990), among others (see special issue of *Developmental Psychobiology*, 1990, *23(7)*, for debate on both sides of the innateness issue. For the purposes of the present discussion, we will focus on one of those problematic assumptions from Chomsky's claims about the LAD, exemplified in the following quote. The notion it conveys, that the input from the environment is inadequate in itself to directly specify the grammar of a language to a learner, characterizes a broader epistemological paradox of historical concern to epistemologists and perception theorists.

> "The native speaker has acquired a grammar on the basis of very restricted and degenerate evidence; the grammar has empirical consequences that extend far beyond the evidence. At one level, the phenomena with which the grammar deals are explained by the rules of the grammar itself and the interaction of these rules. At a deeper level, these same phenomena are explained by the principles that determine the selection of the grammar on the basis of the restricted and degenerate evidence available to the person who has acquired knowledge of the language, who has constructed for himself this particular grammar." Chomsky, 1972, p. 27)

Chomsky asserts in numerous places in his writings that the spoken input from the language environment provides inadequate information about the underlying grammar of the language for the child to apprehend that grammar directly. As the argument goes, each utterance of adult models offer the young child only an incomplete glimpse of the grammar of the language; some utterances are even ungrammatical. Moreover, caregivers generally fail to provide the sort of negative evidence that would unequivocally refute any incorrect hypotheses the child might entertain about the grammar of the language (e.g., Marcus, Pinker, Ullman, Hollander, Rosen, & Fei, 1992). In short, the input is a sample of utterances that, individually, are incomplete (and consequently, sometimes ambiguous) reflections of the underlying grammatical system, and that, collectively, presents but a tiny subset of the infinite grammatically acceptable sentences that a native speaker-hearer could automatically understand and produce.

Thus, the input utterances are taken to be informationally inadequate to specify the grammar completely and uniquely. Therefore, the reasoning proceeds, the child must innately possess a specialized device to construct a model of the grammar and test hypotheses against this input. Because this sort of data base has the potential to permit a large number of logically possible alternative descriptions of a grammar, innate constraints on the forms of permissible grammars are posited to be built into the LAD. Although these arguments have been developed primarily to account for acquisition of syntactic processes, it is presumed that phonology is subject to the same general principles as syntax. The surface phonetic input inadequately specifies the underlying phonological system, therefore phonological acquisition must depend on innate mechanisms. In the remainder of the current discussion, comments about the acquisition of grammar refer primarily to the phonological component (see Dent, 1990, re: similar criticisms of nativist claims about semantic and syntactic development).

Here is the crux of the paradox: The grammar of a language, including its phonology, must be shared sufficiently well by the members of the language community for them to understand each other's utterances. Chomsky's argument is that the child cannot get the grammar directly from the inadequate evidence provided by adult utterances, and so must use innate linguistic mechanisms to determine the grammar. But how can a shared grammar be developed in this way, individual mind by individual mind, based on inadequate input? How could such private grammars ever be verified, given the presumed inadequacy of the utterances[3] which are the only direct evidence that speaker-hearers can present to one another? How could those private grammars become mutually adjusted so that their users would be speaker-hearers of the same language?[4]

Chomsky's solution apparently is that the basis of this mutual adjustment is the innate endowment of linguistic concepts in the universal grammar that all humans share. Those innate concepts are employed to generate and test hypotheses about the grammar of a language against the primary linguistic data each child receives. However, as Chomsky acknowledged, a given set of primary linguistic data usually will support multiple solutions. To keep this problem from getting out of hand, he proposed that the number of potential solutions is limited by innate constraints on permissible grammatical forms. Nonetheless, multiple grammatical hypotheses are still to be expected; the language learner must select the "best" of the possible grammatical hypotheses generated to account for the observed data. Evaluation criteria for choosing the best among a set of possible solutions generally rely on concepts such as elegance or simplicity, which can be notoriously difficult to define and reach consensus on (see Anderson, 1985; cf. Jeffreys & Berger, 1992). Again, the handling of this problem is attributed to innate mechanisms—the requisite linguistic evaluation criteria are part of the LAD. But the difficulties of this line of explanation remain, compounded by the fact that the linguistic data set each individual receives will be different in particulars from that received by each other individual, even within the same community. Given this fact, how would the individual children of a language community end up generating and selecting the same, or similar-enough,[5] grammars?

All normal children, and many who are exceptional in some way, acquire the language spoken to them within a few short years. If the similarities among their disparate input sets are sufficient for children of a language community to select the same (or quite highly overlapping) "most elegant" solutions from among the various alternative grammars that each one privately generates, then surely this must mean that the input from adults provides robust and consistent, rather than inadequate, evidence about the grammar of the language. Indeed, if this be the case, why must the children construct their own private grammars at all? Why not learn the grammar directly from the patterning of the publicly available information in utterances, i.e., learn the phonological system directly from the surface phonetic patterning of utterances?

The problems just summarized reduce to the philosophical paradox inherent in indirect theories of perception. The paradox has been recognized historically even by proponents of indirect theories. Specifically, it is that if inputs convey inadequate veridical information about the world, then we cannot directly know the outer world. The notion that we must know the world only indirectly, through deduction and interpretation of inadequate input, comes down to a claim that we can perceive in the world only what we already know is there to be perceived. This is, of course, the reasoning behind the standard nativist claim for innate knowledge. And as James Gibson (1979) argued, it is circular reasoning.

> "Note that categories cannot become established until enough items have been classified but that items cannot be classified until categories have been established. It is this difficulty, for one, that compels some theorists to suppose that classification is a priori and that people and animals have innate or instinctive knowledge of the world. The error lies...in assuming that either innate ideas or acquired ideas must be applied to bare sensory inputs for perceiving to occur.... Knowledge of the world cannot be explained by supposing that knowledge of the world already exists."(J. J. Gibson, 1979, p. 252-253)

The claim for innate ideas would also seem to be at odds with the basic evolutionary principle of natural selection, dependent as that principle is on the organism's fit to an ecological niche. That is, a species' survival is optimized when its physical structure and behaviors are well-suited to those veridical properties of its world that are relevant to satisfying its procreative and survival needs. I would argue that, as applicable as these concerns are for indirect theories of perception of the physical world, they apply equally to Chomsky's nativist model for acquisition of the phonological grammar of a language. In particular, they are directly relevant to the assumptions that model makes about indirect perception of phonetic patterns in speech.

A fundamental problem of the indirect perception view is that it conceives of input to the perceiver from the world as a series of instantaneous collections of stimulus features which impinge on the special sensory organs (i.e., eyes, ears, nose), and which inadequately specify their dynamic and substantive sources in the world. Like snapshots, these inputs individually have no extension in time or space. A somewhat analogous view can be found in the nativist linguistic assumptions about the language input to the child, which could be characterized as "sound-bites" of language—individual utterances each of which can provide only partial evidence about the underlying grammar, including its phonological component. According to indirect perception theories, because the stimulus cues are impoverished with respect to real-world events and objects, the perceiver presumably must use additional mechanisms of brain and/or mind to further process the sensory inputs, deduce what their sources must have been, draw inferences, develop memorial associations, *etc.*, in order to mentally construct an indirect representation of the world. But how could such mechanisms ever have evolved, given that the presumed inadequacy of the input would make it impossible for their outputs ever to be verified *vis à vis* the real world?

It was in response to these and other sorts of concerns about indirect theories of perception and perception-dependent knowledge in general that the Gibsons formulated an alternative, ecological approach to perception and perceptual learning (E. Gibson, 1969; J. Gibson, 1966, 1979). They argued that all animals, for the sake of their survival, must know the world *directly* from information available in stimulation.

The direct realism alternative: Gibsons' ecological theory of perception

The ecological theory of perception represents the opposite philosophical extreme from the nativist assumptions of Chomsky's theory. The philosophical stance taken by the Gibson's ecological theory of perception is that of *direct realism*, as opposed to indirect or innate knowledge. As the quote below illustrates, ecological theory assumes that stimulation is structured and dynamic, extending over time and space, and that it is directly detected rather than being "interpreted" by innate knowledge, computation, inference, stored memories, or arbitrary associations.

> "The evidence...shows that the available stimulation surrounding an organism has structure, both simultaneous and successive, and that this structure depends on sources in the outer environment. If the invariants of this structure can be registered by a perceptual system, the constants of neural input will correspond to the constants of stimulus energy, although the one will not copy the other. But then meaningful information can be said to exist inside the nervous system as well as outside. The brain is relieved of the necessity of constructing such information by *any* process—innate rational powers, (theoretical nativism), the storehouse of memory (empiricism), or form-fields (Gestalt theory). The brain can be treated as the highest of several centers of the nervous system governing the perceptual systems. Instead of postulating that the brain constructs information from the input of a sensory nerve, we can suppose that the centers of the nervous system, including the brain, resonate to information." (J. J. Gibson, 1966, p. 267).

As this passage indicates, information about the external world—about distal events, surfaces, and objects—is assumed to be directly picked up from stimulation, by integrated perceptual systems. To illustrate the perceptual system concept, the retina of the eye does not gather visual information by working in isolation. Rather, it is an integral part of the perceptual system for seeing: two movable eyes fixed in a head, which is attached to a body that can move to shift location and orientation of the viewer with respect to the external spatial layout; these components are neurally integrated with one another and with higher centers in the brain. Thus, the perceptual systems are assumed to have evolved to permit active, physical exploration of the world in the service of gathering and disambiguating distal information.

Thus, the ecological approach, like the linguistic nativist approach espoused by Chomsky, is concerned with biological specialization. However, the two views differ dramatically in their assumptions about the nature of biological specializations—the information they handle, the way they work, and the forces behind their evolution. According to ecological theory the biologically specialized perceptual systems have evolved, and continue to function, for the pick-up of veridical information from the world. This view admits the possibility of perceptual systems being specialized for pick-up of information about specific types of distal objects or events, such as the information in speech that specifies the configuration and movements of the vocal tract producing the signal (see Best, 1984, 1993, in press, a, b). Such specializations may be abstractly analogous to that of the human hands for grasping and manipulating objects, and the complementary perceptual ability to detect the graspability and manipulability of distal objects. Evidence for primitive components of the latter abilities, and of their responsiveness to the physical properties of distal objects (size, distance, speed of movement) is found quite early in development (e.g., von Hofsten, 1980). As for the pick-up of distal articulatory information in the speech signal, Gibson summarized in general terms how and why this should be possible (see also Best, 1984, in press a, b; Fowler, 1986, 1989, 1991):

> "An articulated utterance is a source of a vibratory field in the air. The source is biologically "physical" and the vibration is acoustically "physical." The vibration is a potential stimulus, becoming effective when a listener is within range of the vibratory field. The listener then *perceives* the articulation because the invariants of vibration correspond to those of articulation. In this theory of speech perception, the units and parts of speech are present both in the mouth of the speaker and in the air between the speaker and the listener. Phonemes are in the air. They can be considered physically real if the higher-order invariants of sound waves are admitted into the realm of physics." (J. J. Gibson, 1966, p. 94)

The direct realist philosophy assumes that information from the world is a rich multimodal flow of temporally and spatially distributed energy patterns that are lawfully and systematically shaped by distal events and objects. The systematic structure in this information flow is picked up by perceptual systems—extracted, detected, discovered—through active, physical exploration of the events, surfaces and objects that shape the energy flow. By shifting position and orientation with respect to the objects and the spatial layout, as well as by moving and manipulating objects, the perceiver produces changes in the flow of stimulation that are systematically influenced by the exploratory actions in ways that provide rich, direct, veridical information about the distal sources of stimulation. As a result of this active exploratory behavior of the perceptual systems, the perceiver becomes better attuned, with increases in experience, to the invariants in stimulation that specify the defining characteristics of specific events, the persisting identity of particular objects, and the higher-order commonalities shared by similar events or by similar objects.

The transformational invariants of an event are those properties of the energy flow that remain constant across the participation of different objects in that event. For example, the transformational invariant of repetitive rotation about an axis specifies the same event of spinning whether a top is spinning on a surface, an amusement park "anti-gravity" ride is spinning to produce centrifugal force, or the wheels of a car are rotating on their axles. The structural invariants of spherical shape and elastically deformable solid specify an identity relation—the same baseball across the events of rolling, throwing, bouncing, and juggling. Invariants can also specify similarity relations among objects or events. The more abstract invariant of a convexly-curved plane characterizes the primary similarity among the outer surface of an eyeglass lens, the dome of an enclosed sports arena, and the silhouette of an old Volkswagen "beetle." And although the following do not reflect literally the same event, they involve abstractly similar curvilinear movement transformations: the slithery, winding progression of a snake, the sinewy movements of a traditional Thai dance, and the wave-like motion of tall grass rippling in a breeze (for further discussion of structural and transformational invariants, see Shaw, McIntyre, & Mace, 1974). Experience-dependent changes in attunement to such invariants occur through perceptual learning.

The ecological perspective has concerned itself primarily with general perceptual principles rather than with linguistically specialized mechanisms. However, I believe it is eminently applicable to children's learning of the sound pattern of their native language, and to the concomitant effect of this learning on the perception of non-native sounds and contrasts. If we take an ecological view on the realm of language, the spoken input available to the young child is a flow of many utterances, occurring multimodally within a rich behavioral context that extends over time and people. The flow of this linguistic and social stimulation, extending as it does over time and speakers, should reveal regularities or invariants across utterances that the infant comes to recognize as the sound-organizing principles of the phonology of the language (e.g., Best, in press, a).

I have taken the ecological perspective to account for how experience with the ambient language comes to influence the infant's perception of non-native speech contrasts. To do so, I will apply this perspective to linguistic insights about the sound structure of languages, which should form the basis for the child's developing recognition of the relations between the phonetic properties of speech and the phonological organization of the grammar of his or her native language. For the purposes of this chapter, we are particularly interested in how ecological principles apply to perceptual learning, specifically with respect to infants' and young children's perception of the sound pattern of their native language. Therefore, we will turn next to examine in greater depth the ecological approach to perceptual learning.

The ecological perspective on perceptual learning

Two quotes exemplify the ecological viewpoint on perceptual learning, the first from James Gibson's (1979) book *The ecological approach to visual perception*, the second from Eleanor Gibson's address on "Perceptual development and the reduction of uncertainty" at the 18th International Congress on Psychology in Moscow.

> "The perceiving of the world begins with the pickup of invariants.... [T]he theory of information pickup...needs to explain learning, that is, the improvement of perceiving with practice and education of attention.... The state of a perceptual system is altered when it is attuned to information of a certain sort. The system has become sensitized. Differences are noticed that were previously not noticed. Features become distinctive that were formerly vague." (J. J. Gibson, 1979, p.254)
>
> "Discrimination learning proceeds...by discovering distinctive features of objects and invariants of events in stimulation.... The effective stimulus which active and educated perception picks out is a reduced stimulus. It is extracted, filtered out, whereas other stimulus information which has no utility for differentiation is ignored by the educated attention." (E. J. Gibson, 1966, pp. 10-15)

When a perceptual system becomes attuned to a particular type of information, it becomes altered by experience. The claim is that the attuned perceiver is more quickly and efficiently able to pick up from the flow of stimulation just that information to which the perceptual system has become sensitized, as opposed to, perhaps, simply increasing the speed of a cognitive search through mental space. This sensitization of the perceptual system entails detection of critical distinctions among objects or events that had previously gone unnoticed. What it is suggested by perceptual learning, then, is an optimization and economization of pickup or extraction of critically distinctive properties. Perceptual learning is probably more readily apparent for detecting abstract, higher-order invariants (such as the curvilinear movement invariant described earlier) than for detecting the simple, lower-order invariants to which perceptual systems are innately tuned even very early in life (e.g., basic color categories: Bornstein, 1979).

These principles have been more completely drawn out by Eleanor Gibson in her numerous writings on perceptual learning (e.g., E. Gibson, 1963, 1966, 1969, 1977, 1988; E. Gibson & J. Gibson, 1972; J. Gibson & E. Gibson, 1955). As her opening quote indicates, perceptual learning leads to improved discrimination, but this does not mean simply the discrimination of smaller and finer stimulus differences, hence of always increasing numbers of individual stimuli. Instead, perceptual learning entails the discovery, for specific purposes, of the *critically* distinctive features of objects and invariants of events in stimulation. It involves the education of attention for most efficient detection of the most telling differences among objects and events that are of importance to the perceiver. As she has argued, the utility that critical distinguishing features and invariants of events have for the perceptual learner is that they reduce uncertainty among choices in a world that otherwise presents too much, rather than too little, information. Educated attention, i.e., a perceptual system that is attuned to certain types of information, picks up *reduced* stimulus information, which is selected, extracted, or filtered out from the larger flow specifically because of its ability to critically differentiate things that are of interest or usefulness to the perceiver. Other stimulus information that does not serve this purpose of utility is ignored, i.e., not picked up.

This account leaves open the possibility for re-education of perception, because the undetected information is still available in stimulation. Stimulus information that is irrelevant for well-used distinctions, and therefore has been systematically ignored, could later prove important for other new distinctions. It is conceivable, perhaps even likely, that having first learned to economize information pickup by overlooking certain information as irrelevant (or by perceiving it as equivalent to some other pattern of information) may make it more difficult to re-learn to attend to it later than would be the case for a novice learning to attend to the same information for the first time. Ecological theory has not directly addressed these possibilities. However, they are relevant for understanding whether and to what extent second language learners may learn to detect non-native phonetic distinctions that are not utilized in their native language, and in what way this may be affected by varying degrees of experience with the native language.

Indeed, the Gibsons did not address speech perception in great detail in their primary accounts of the ecological approach to perceptual learning (cf. E. Gibson & J. Gibson, 1972), although Eleanor Gibson did address certain aspects of language in her research on reading development (e.g., E. Gibson, 1971). The ecological view on perceptual learning has primarily addressed the *general* issues of how perception is shaped by experience. Perceptual learning entails the discovery of invariants in stimulation that reveal the structural and functional properties of the source objects and events. Often, these invariants are hierarchically nested in complex events, so that higher-order invariants may depend on, or be derivatives of, lower-order invariants. Discovery of certain higher-order invariants may thus be possible only once the perceiver has learned which of the lower-order invariants are critical to the distinction and which are not. Perhaps, for some distinctions, there may even be several levels of lower invariants supporting the discovery of a higher-order invariant.

Spoken language provides an excellent example of the sort of complex organization in which higher-order invariants, such as those that specify syntactic principles, may not be detectable until the perceiver has learned to pick up certain distinctive information at lower levels, such as the critical differences in the phonetic patterns of similar-sounding but meaningfully different words. For the infant, then, learning the sound pattern of the native language is the quintessential task of perceptual learning, i.e., discovering the multiple levels of invariant principles by which the stimulus flow is patterned.

The ecological premise is that the complex, nested hierarchy of linguistic organization, including phonological patterning, exists in the infant's language environment. It is all there, that is, if we consider the available language stimulation to span the history of utterances the infant hears, along with the rich behavioral contexts in which those utterances occur. The flow of spoken utterances in context provides the infant a window on the patterning of the ambient language. This is the flow of stimulation from which infants must learn to recognize and abstract the invariants that specify all levels of linguistic structure. Of course, the infant is not initially able to detect or abstract from that flow the invariant properties specifying most of the levels of linguistic organization summarized above. In fact, the only level of the available information that the infant is likely to be able to detect initially is the surface phonetic information. And it is necessarily from among those phonetic details that the infant must learn to recognize the higher-order invariant patterns that specify words, syntax, morphology, and in particular, phonology.

Thus, the ecological view is that utterances provide a rich flow of information about dynamic speech events which extend over time, and that through perceptual learning the individual becomes attuned to various levels of invariant structure available in that flow. This view suggests a radical departure from the standard assumption of discrete, timeless features and calls instead for a model of phonetics and phonology in which the crucial dynamic attributes of events in the speech world are integral to the model. The ecological perspective has begun to offer alternative insights and evidence both about the phonetic details of speech production (Fowler, Rubin, Remez, & Turvey, 1980; Kelso, Saltzman, & Tuller, 1986; Saltzman & Munhall, 1989), and also about its phonological organization (Browman & Goldstein, 1986, 1989, 1990a, b, c, 1992a; Fowler, 1980; Goldstein & Browman, 1986). The latter work has offered an articulatory gestural model of phonology, which we will examine next as the basis for an ecological, perceptual learning account of language-specific effects on the perception of non-native phonetic contrasts. The following summary is based on the works of Browman and Goldstein cited above.

Gestural phonology

The tenets of gestural phonology are grounded in the spatiotemporal organization of articulatory gestures in speech, which are themselves grounded in the biomechanical organization of the human vocal tract. Rather than assuming abstract and timeless phonetic features as the atoms or primitives from which phonological representations are built, the gestural model assumes that the phonological primitives are articulatory gestures, the coordinated actions of vocal tract articulators. The model organizes these gestural features within the framework of a hierarchical ***articulatory geometry*** based on the anatomical relations among the articulators involved in speech. The vocal tract is comprised of three relatively independent articulatory systems that are represented as separate nodes within the articulatory geometry: the glottal system (vocal cords), the nasal system (the velum, the valve that permits or prohibits air flow through the nasal cavity), and the oral system, which includes the lips and the tongue as separate subsystems. There is an additional subordinate level in the tongue subsystem: tongue tip versus tongue body, whose actions are differentiated by different intrinsic and extrinsic muscles of the tongue. This hierarchically organized set of articulators functions within the confines of the walls of the vocal tract, which is structured basically as a bent tube of varying diameter, optionally connected to a second side tube (nasal cavity) via the open velum. The coordinated actions of the articulators can cause constrictions at various locations (place of articulation) along the vocal tract (e.g., dental, alveolar, velar, etc.) (see Figure 1 for additional places of articulation). Each place can display several variations in degree of constriction, which determines the manner of the sound produced (complete closure for stop consonants, critical constriction for causing turbulent airflow in fricatives, narrow constriction for some vowels and for approximant consonants such as /w/ and /r/, wide opening for the velum in nasals and the glottis in voiceless sounds). Articulatory geometry is compatible, in many respects, the with nonlinear or autosegmental approaches that have supplanted SPE phonology. Some important distinctions must be noted, however, between the two approaches. Specifically, gestural phonology posits phonological elements to be gestures defined by a set of dynamic equations describing the movement of articulators over space and time, rather than a specification of abstract, timeless phonetic features. To illustrate, the equation set for the syllable *ma* describes a velum opening gesture and lip closing gesture which begin simultaneously and reach their peaks synchronously to produce the /m/, and a slower, less extreme tongue body gesture to narrow the pharynx (upper throat) for the "ah" vowel, which begins synchronous with the other two gestures but peaks later and lasts longer.

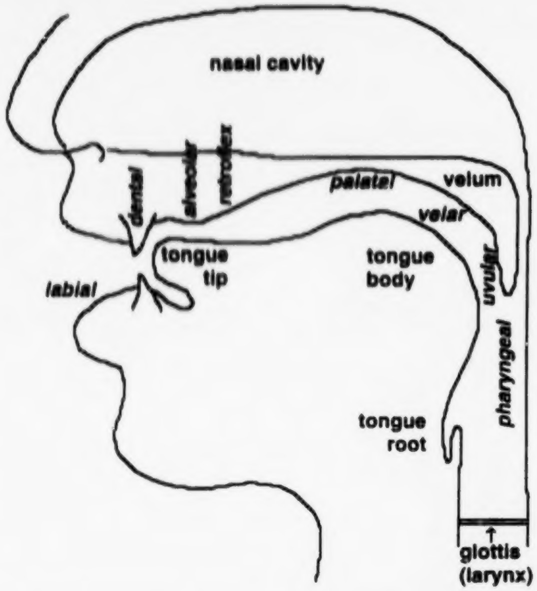


Figure 1. Schematic lateral view of vocal tract, with major articulators labeled and the nasal cavity identified. Many of the common places of articulation, or locations of articulatory constrictions, are indicated in italics.

Thus, articulatory geometry is closely related to the anatomical structures and movement patterns of the vocal tract. This way, in the gestural model the phonological primitives and their physical instantiations derive from a single domain grounded in the spatiotemporal properties of real articulatory events. Because of this, phonological representations can specify the relative timing, or phasing, of one articulatory gesture relative to another. For example, the Canadian French versus continental French difference in vowel nasalization that was mentioned earlier (van Reenin, 1982) can be specified dynamically as a difference in the relative timing, or ***phasing***, between the onset of velum lowering for nazalization and the peak of tongue movement for the vowel. This characterization departs critically from the phonetics-phonology relationship held by classic SPE phonology and by nonlinear phonologies, neither of which can phonologically represent the dialectal difference phonologically, even though the nasalization difference appears to be part of the language-specific grammar in the two dialects. This representational inability occurs for the latter two views because they posit that phonetic and phonological information exists in two divergent, informationally incompatible domains, one physical (actual articulations) and the other only mental (underlying phonological representations).

In gestural phonology, the dynamical specifications of articulatory gestures describe change over time in particular vocal tract variables and their associated articulators (e.g., location and degree of a constriction by the tongue tip or tongue body somewhere along the vocal tract tube; opening of the nasal tube by movement of the velum). The model assumes that articulator motion is governed by dynamic principles of spring-like physical systems,[6] in which the values of several parameters of the tract variable(s) are specified: mass, stiffness, damping, rest position, instantaneous position, acceleration and velocity. All tract variables are assumed to have a resting, or default, setting. The resting state is not, of course, specified as a gesture; gestures are active articulatory movements away from the resting state. A given gesture is a particular transformation of a tract variable (e.g., complete closure of the lips) that remains invariant across different contexts, speaking rates and styles, and speakers. There may also be variation in the exact articulators or coordinations among articulators that are used to achieve essentially identical gestural goals. For example, bilabial closure may be achieved by moving only the lips and keeping the jaw angle constant, or by keeping the lips immobile and changing only the jaw position to bring the lips closer together (see Abbs & Gracco, 1984). Therefore, the dynamical description of a particular gesture defines a family of articulatory trajectories that all achieve the same gestural target of a particular degree of constriction at a particular location along the vocal tract tube.

Some phonological elements are composed of only a single gesture, whereas others involve a specific pattern of coordination between two or more individual gestures. Coordinations among two or more gestures are called ***gestural constellations***. Let us illustrate the difference with the /p/-/b/ contrast, which in classic phonological description share the phonetic features [+anterior], [-continuant] and [-sonorant], and are distinguished only on the feature [+/- voice]. But in gestural description, the voiced stop /b/ in *gabbing* involves only a single bilabial closure gesture (complete closure and release of constriction at the lips). The state of the glottis, or opening between the vocal folds, is maintained in the default adducted position (critical constriction rather than tightly closed) and produces voicing throughout the word. In other words, there is no active glottal gesture just for the /b/. In contrast, the cognate voiceless stop /p/ in *gapping* involves two gestures which must be correctly phased relative to each other. Specifically, the bilabial closure must co-occur with an active glottal opening gesture, which prevents voicing and instead permits turbulent airflow (i.e., aspiration noise) through the vocal folds. The peak opening of the glottis coincides with release of the bilabial constriction; the glottis returns to its default state (vocal folds together for voicing) after bilabial release. The /p/ example illustrates a gestural constellation that corresponds to the segmental level of traditional phonology. But gestural constellations may also describe articulatory coordination at the level of syllables, words, prosodic phrases, etc. Analogous to nonlinear phonological approaches, these nonsegmental levels of linguistic organization among gestures are specified for different articulatory tiers, such as those representing syllable structure and stress units. However, neither gestures nor constellations bear a one-to-one relationship either to segments or to classic phonetic features.

Because gestures are defined by a dynamical pattern of articulatory movements, each gesture has both an intrinsic spatial aspect and an intrinsic temporal aspect. This grounding in the physical properties of events over time departs qualitatively from the classic and the nonlinear views of static, dimensionless phonetic features. In gestural phonology, the phasing principles among the gestures in a given utterance are represented in both their spatial and temporal relations in a ***gestural score***. To illustrate, a schematic gestural score for the word *mob* ([mab]) is shown in Figure 2. The abscissa represents the time line of the utterance, the ordinate represents the tiers in the articulatory geometry that are needed to display the critical gestures involved in that particular word. The rectangular boxes represent the temporal extent during which given gestures are active for their corresponding articulatory tiers, or articulatory sets (e.g., tongue tip, tongue body, etc.). Inside each activation interval box, the degree of constriction achieved in the gesture and its specific location along the vocal tract are denoted. An American English utterance of *mob* begins as was described earlier for the syllable *ma*. The pharyngeal gesture for the vowel ("ah") extends into the final bilabial closure that corresponds to /b/.

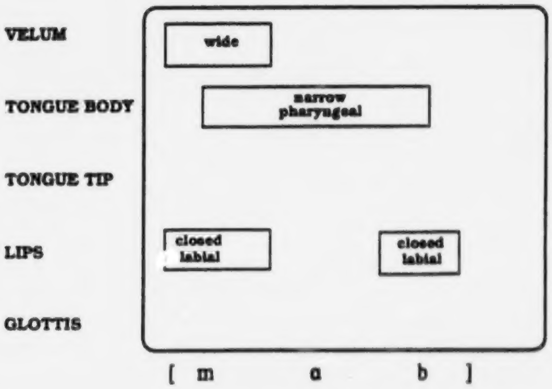

VELUM
TONGUE BODY
TONGUE TIP
LIPS
GLOTTIS
wide
narrow pharyngeal
closed labial
closed labial
[m ɑ b]

Figure 2. Schematic gestural score for the word <mob> [mɑb] using box notation to indicate activation intervals for gestures and phasing among gestures.

Thus far, the gestural phonology approach has been applied in detail primarily to American English alone, but it can be extended (and in some cases has been) to suggest gestural characterizations of certain similarities and differences between the gestural constellations for some non-native phonetic contrasts and contrasts found in the English phonological system. A few cross-language comparisons will be offered here as illustrations. However, we must bear in mind an important caveat from Browman and Goldstein (1992b), that any proposed gestural analysis is obviously incomplete and speculative in the absence of hard data on the actual gestural processes involved in the utterances being considered. The comparisons here are based on currently available phonetic, acoustic, and physiological descriptions for the phonological contrasts involved. But the schematic gestural scores offered are necessarily speculative because of the incompleteness of actual gestural evidence, especially with respect to temporal extent and precise phasing of gestures.

Figure 3 shows the Hindi dental-retroflex contrast [ɖɑ]-[ɖɑ] and English [dɑ], which is gesturally most similar to both Hindi patterns. The schematized Hindi gestural scores and the English one are essentially the same except that the Hindi constriction locations are just anterior and just posterior, respectively, to the English alveolar location. Recall also that English does have context-conditioned dental and retroflex allophones of /d/, but not in the context of an isolated [dɑ]. Schematic gestural scores for the Zulu aspirated versus ejective velar stops [kʰa]-[k'a] are compared to the correspondingly most similar English gestural constellation, that for [kʰa], in Figure 4. In this case, the Zulu aspirated token is virtually identical to the English one, whereas the ejective token deviates from it in the constriction degree of the glottal gesture, which is closed rather than wide, producing silence rather than aspiration prior to the onset of voicing for the vowel. A different type of Zulu contrast is between voiced and voiceless lateral fricatives. These gestural constellations are produced with essentially the same alveolar tongue tip closure and uvular tongue body narrowing as in English /l/. They differ, however, in employing a smaller constriction degree along the two sides of the tongue (against the upper lateral teeth) than for /l/. Instead, the lateral constriction is critical, producing airflow turbulence analogous to that at

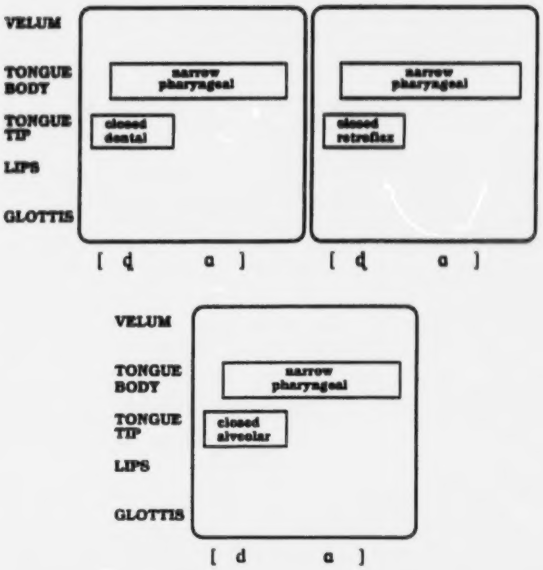

VELUM
TONGUE BODY
TONGUE TIP
LIPS
GLOTTIS
narrow pharyngeal
closed dental
[ɖ ɑ]
narrow pharyngeal
closed retroflex
[ɖ ɑ]
VELUM
TONGUE BODY
TONGUE TIP
LIPS
GLOTTIS
narrow pharyngeal
closed alveolar
[d ɑ]

Figure 3. Schematic gestural scores for the Hindi dental-retroflex /ɖɑ/-/ɖɑ/ contrast (top panels) and English /dɑ/ (bottom panel).

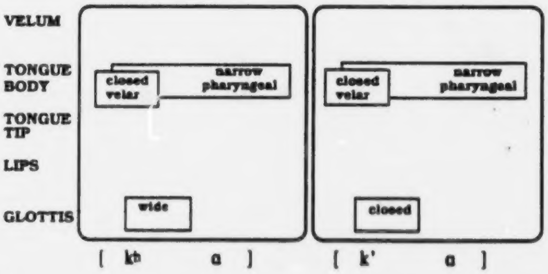

VELUM
TONGUE BODY
TONGUE TIP
LIPS
GLOTTIS
closed velar
narrow pharyngeal
wide
[kʰ ɑ]
closed velar
narrow pharyngeal
closed
[k' ɑ]

Figure 4. Schematic gestural scores for the English and Zulu voiceless velar stop /kʰɑ/ (left) and for the Zulu ejective velar stop /k'ɑ/ (right).

the tongue tip for fricatives such as English /z/ or "zh" or voiced "th" (in *that*) versus /s/ or "sh" or voiceless "th" (in *think*). Thus, the Zulu lateral fricatives gesturally resemble both the liquid /l/ and the voiced-voiceless fricative distinctions of English that involve tongue tip constrictions at anterior locations. Larger English gestural constellations (multi-segmental) that may approximate the patterns found in the lateral fricatives include /zl/-/sl/ (*paisley*, *slow*), "zhl"-"shl" (*rougeless*, *Ashley*), or voiced vs. voiceless "thl" (*blithely*, *breathless*). Finally, the Zulu alveolar versus lateral click consonants incorporate gestural constellations that are quite dissimilar from any in English. Both have full closures at two locations, alveolar (tongue tip) and velar (tongue body). A vacuum is created in the intervening zone by drawing the tip or one side of the tongue downward until the suction is released. In syllabic context, this is followed immediately by release of the velar closure. The double closure plus suction release does not closely resemble any English gestural constellation.

Gestural phonology can also account parsimoniously for a wide variety of phonological phenomena within its articulatory framework, using gestural primitives that have intrinsic temporal and spatial dimensions, unlike static, dimensionless phonetic features. In most cases these gestural accounts are backed by speech production data. For example, minimal contrasts are two gestural constellations that are identical except for a critical difference in constriction location (e.g., /b/ vs. /d/) or constriction degree (e.g., /b/ vs. /w/) in the oral tier of the articulatory geometry, presence/absence of a gesture of the velum (e.g., /ma/ vs. /b/) or glottis (/p/ vs. /b/), etc. The tube geometry of the vocal tract also appears to account straightforwardly for certain natural classes, i.e., groupings of different types of phonetic categories that nonetheless participate together in widespread phonological processes. To illustrate, nasals, liquids (/r/, /l/) and vowels form the class defined traditionally by the [+sonorant] feature, which has been difficult to define objectively. In gestural phonology, these phonetic types share the simple gestural similarity that they all maintain one of the two vocal tract pathways (oral, nasal) wide open for outward airflow (Browman & Goldstein, 1989). Many allophonic variants can be explained as the overlapping of adjacent gestures, or coarticulation, as in the dental allophone for /n/ in *ten themes*, which results from overlapping of the wide velum for /n/ and the dental location of the tongue tip for "th" (Browman & Goldstein, 1989). Analogously, gestural overlap can account for certain cases of phonological assimilation, as when the /n/ in *seven plus* assimilates to /m/ in casual speech. The feature-based rule is that the labial feature of the /p/ spreads forward to the /n/. The gestural explanation is that the bilabial closure gesture of the /p/ overlaps the velum opening gesture for the /n/, thus "hiding" the aerodynamic evidence of the alveolar tongue gesture for /n/ and producing the bilabial nasal /m/ (Browman & Goldstein, 1989). Cases of phonological deletion can be handled likewise from a gestural perspective. For example, feature-based approaches posit a deletion rule whereby the final /t/ of the first word in *perfect memory* gets deleted, but in gestural terms it is simply the case that the alveolar /t/ gesture gets hidden by overlap with the /m/ of *memory* (Browman & Goldstein, 1989, 1990c). Gestural overlap can even account for the insertion of an additional segment between other segments, called epenthesis. As an illustration, *something* is often pronounced in American English with a /p/ between the /m/ and the "th," leading feature-based accounts to invoke an insertion rule. But the /p/ arises gesturally from the overlap of the bilabial closure gesture for the /m/ and the glottal opening gesture for the following "th" (Browman & Goldstein, 1990c). The phenomenon of metathesis, in which the sequential order of segments becomes reversed by some phonological process, has been particularly vexing for generating feature-based rules that are powerful enough to describe the phenomenon but not so overly powerful as to generate many non-occurring reversals. Such ordering reversals often occur in speech errors, as when the rapid production of *Bob flew by Bligh Bay* comes out as *Blob foo by Bligh Bay*. A gestural analysis of tongue movement for the /l/s in these utterances reveals evidence of the temporal "sliding" or overlap of the tongue tip constriction gesture with those preceding it in the represented sequence, causing both overt and covert speech errors (Browman & Goldstein, 1992a).

The gestural phonology model has received some criticism from nonlinear phonologists, as well as some praise. On the positive side, some phonologists acknowledge that placing articulatory constraints on phonological processes is advantageous (see also Archangeli, 1988; Archangeli & Pulleyblank, in press), especially with respect to better delineation of the relation between phonology and phonetics (e.g., Clements, 1992; Pierre-Humbert & Pierre-Humbert, 1990). By and large, the criticisms reflect two underlying observations:

1) gestural phonology rejects static, timeless phonological features that differ in kind from physical, phonetic realizations; 2) it does not invoke abstract cognitive rules about phonological representations (e.g., Pierrehumbert, 1990; Pierrehumbert & Pierrehumbert, 1990; Steriade, 1990). In other words, gestural phonology rejects two central tenets held by both SPE and nonlinear phonologies. These criticisms also suggest some partial misunderstanding of gestural phonology. The model does include discrete, or categorical, elements at the phonological level of the task dynamics used to generate gestures (Browman & Goldstein, 1992b). Moreover, it does distinguish between phonological and phonetic levels of representation, but views them as macroscopic versus microscopic descriptions of the same dynamic, physical domain of speech events (Browman & Goldstein, 1990a; see also Ohala, 1990). This brings us back to the central claims of the ecological approach, which assumes that perception must be grounded in physical reality. On that note, let us return to the issue of how the physical properties of native speech are perceived by the adult and learned by the child.

The ecological approach to perceptual learning of speech

All of the phonological approaches discussed, including gestural phonology, have taken their task to be the generation of a physical phonetic output from the more abstract phonological component of the grammar. But we began with, and now return to, the opposite process—how a perceiver, particularly a young learner, gets from the phonetic surface to the phonological structure via perception. Specifically, the chapter began with the question of how experience with one's native language comes to affect one's perception of non-native speech sounds and contrasts from unfamiliar languages. Phonology has provided little guidance here. Although Chomsky and Halle stated in SPE that a listener who knows the language being spoken will hear the phonetic shapes predicted by the phonology, it is unclear how they would expect the phonology to handle discrepancies between the phonetic features in a non-native sound and the feature matrices defined by the phonological system of the listener's language. Indeed, how would it even handle perception of corrupted native speech (e.g., foreign accented or disordered speech), or the phonetic patterns of an unfamiliar dialect? Nonlinear phonological approaches don't help much, as they also have devoted minimal attention to theoretical issues in perception. And gestural phonology, the youngest of the approaches, has also focused the majority of effort on production. Moreover, none of these phonological approaches has given any depth of consideration to how infants and young children perceptually learn about the phonological structure of their native language.

To address these issues, we return to the direct realist view of speech perception based on the Gibsons' ecological theory of perception. This view assumes that listeners perceive information in speech about the distal articulatory gestures that shaped the phonetic patterns (Best, 1984, 1993, in press, a, b; Fowler, 1986, 1989, 1991;). Because it assumes that phonological processes derive from the same physical, dynamic domain as the phonetic details of actual utterances, gestural phonology lends itself to an ecological perspective on cross-language influences in perception, as well as on how the infant learns the phonological properties of the native language. Articulatory gestures would provide a common metric for both perception and production of speech. The interrelation of perception and production is central to both speech imitation and language acquisition.

The direct realist view posits that perceivers recover information from speech, and from other sound-producing events, about the distal structures and events that produced the sounds. This view assumes that information about articulatory gestures is *directly perceived* in speech, as opposed to being the end-product of cognitive processing of the raw acoustic input. The speech signal is shaped by the structure and movements of the vocal tract according to physical laws, as indicated by the earlier quote from James Gibson. Thus, evidence about articulatory gestures is available to perceivers as structured information about the speech events that produced the signal. This view is *not* the same as that of the well-known motor theory of speech perception (e.g., Liberman & Mattingly, 1985), which posits that perceivers refer to the motor control of their own speech in order to perceive the phonetic structure of speech input. The ecological claim is that listeners perceive the speaker's articulatory gestures as such, without referring to their own articulatory commands and, indeed, regardless of whether they can themselves produce similar signals.

That listeners perceive gestural information in speech is supported by cross-modal speech perception research (see also Best, 1993; Studdert-Kennedy, 1993). McGurk (McGurk & MacDonald, 1976) found that when presented with audiovisual syllables in which the synchronized consonants in the two modalities are from different categories, listeners perceive a unified phonetic pattern that is compatible with both modalities, rather than noticing the discrepancy. That is, the two modalities apparently provide evidence about a common, underlying dimension such as articulatory gestural patterns. An alternative argument that the perceptual link between visual and auditory information is learned by association is illogical in the general case, according to the Gibsons' arguments, and has been empirically refuted for the speech perception case by two recent reports. Cross-modal integration does occur for synchronized but discrepant consonants presented auditorily and tactually—blindfolded subjects manually felt the movements of an experimenter's silent lip movements, synchronized with audio recordings—although they had never had such tactile-auditory experience with speech. Yet there was no cross-modal integration for synchronized audio and written syllables, in the face of the subjects' extensive associative experience with the relation between text and speech (Fowler & Dekle, 1991). In another study, young English-learning infants heard repetitive audio presentations of the French lip-rounded vowel /y/, which does not occur in English, synchronous with side-by-side silent videos of English lip-rounded long "oo" and unrounded "ee" (Walton & Bower, 1993). The infants preferentially fixated on the "oo" video when hearing /y/. Given their lack of prior experience with /y/ this could not have been a learned association, but rather suggests detection of the articulatory commonality of lip-rounding across modalities.

More in-depth treatment of the rationale and evidence for the general direct realist approach to speech perception can be found in other reports (e.g., Best, 1984, 1993, in press, a, b; Fowler, 1986, 1989, 1991; Fowler, Best, & McRoberts, 1990 Verbrugge, Rakerd, Fitch, Tuller, & Fowler, 1984). Our concern here is specifically with how infants' and adults' perception may be differently affected by experience with the native language, particularly by its phonological structure. What we perceive in both native and non-native speech appears to depends what we've learned about the native phonology through experience with that language.

Language-specific phonetic-gestural properties and perceptual learning

Recall the basic tenets of perceptual learning according to the ecological perspective—that perceptual systems become attuned by experience to particular types of information; that this involves optimization in the pickup of relevant information; that it entails the discovery of critically distinguishing properties of distal structures and events; and that this is accomplished via perceivers' active search for invariants in the flow of stimulation that most economically specify those crucial properties. Educated attention minimizes uncertainty about objects and events in the world, by selecting or extracting reduced information specifically for its ability to critically differentiate things of interest or usefulness to the perceiver. Earlier it was argued that the identity of objects and events is specified by structural and transformational invariants available in the flow of stimulation over time and space. Moreover, recognition of similarities and differences among things often depends on abstraction of higher-order invariants which depend on prior detection of other, lower-order invariants. As Eleanor Gibson remarked, the critical invariants are generally relational in nature, rather than isolated, independent attributes.

To consider how higher-order relational invariants might be discovered in speech through perceptual learning, I will turn briefly to some central concepts developed in work on an ecological approach to the formation of complex coordinated skills and behaviors (e.g., Kugler, Kelso, & Turvey, 1982; Saltzman & Kelso, 1987; Turvey, 1980; 1990) including speech (Saltzman & Munhall, 1989). The goal of coordination is to maximize the adaptability and flexibility of achieving some goal of action by minimizing the number of separate dimensions that must be directly controlled. As Turvey (e.g., 1980, 1990) and others have argued, this is accomplished by forming task-specific synergies among muscle groups, or ***coordinative structures***. To understand this concept, consider an example commonly cited by ecological researchers—the task of a puppeteer and the way that the construction of her marionette simplifies the control of its movements. By linking the puppet's limbs with strings to a controller bar, the puppeteer obviates the need to move each joint of each limb separately, instead producing coordinated movements among multiple limbs by single movement of the controller. By this means, the many degrees of freedom controlling the joints of the separate limbs have become joined together into a coordinative structure with fewer degrees that must be directly controlled. Research on locomotion indicates that coordinative structures account for the coordination of flexion and extension of each leg joint in proper sequence during the swing of each leg, the alternation between the legs, and the postural adjustments required throughout for maintenance of balance. Coordinative structures show task-specific flexibility in that temporary perturbations result in automatic, immediate compensatory adjustments among the coordinated elements so that the general goal is preserved without requiring numerous command decisions about specific elements.

Saltzman and Munhall (1989) provide logical and empirical evidence that in speech coordinative structures accomplish the gestural goal of forming a constriction of a particular degree at a particular vocal tract location, by harnessing together the specific articulators in ways that automatically compensate for perturbations and contextual variations. The language-specific gestural phasing patterns of Browman and Goldstein's gestural constellations are examples of higher-order coordinative structures in speech. Coordinative structures in motor control can form and re-form, and operate as emergent properties of self-organizing systems (see Madore & Freeman, 1987; Prigogine, 1980; Prigogine & Stengers, 1984; Schöner & Kelso, 1988; Turvey, 1980, 1990). Emergent properties of self-organizing systems, including their sensitivity to initial conditions, have been proposed as the basis for the evolution of maximal dispersion among the elements of language-specific phonological inventories (Lindblom, 1992; Lindblom, Krull, & Stark, 1993; Lindblom, MacNeilage, & Studdert-Kennedy, 1983), as well as for the ontogeny of phonological organization in the child (Mohanan, 1992; Studdert-Kennedy, 1989). The latter proposals point to the importance of viewing the native phonology as an organized system when considering how language-specific experience may affect perception of phonetic patterns that fall outside the native phonological system.

Insights about coordinative structures and self-organizing processes, and about the importance of minimizing the degrees of freedom that must be separately controlled, will serve as useful heuristics for thinking about perceptual learning of phonetic and phonological structure in native speech. Indeed, they are crucial to an ecological approach to the issue, given the direct realist assumption that speech perception entails the pickup of information about the distal articulatory events that produced the signal. The ecological approach assumes that perceivers actively explore the rich flow of multimodal information in spoken utterances for invariant patterns that are of interest or utility to them. Educated perception should therefore actively seek and extract critical features of the coordinative structures responsible for the gestural organization of native speech. These coordinative structures should include language-specific articulatory gestures and constellations of phasing among gestures at all levels in the language—from traditional segments to syllables, words, prosodic phrases, etc. The information detected for the language-specific coordinative structures would be higher-order invariants, consistent with the principle that an attuned perceptual system optimizes information pickup by extracting a reduced stimulus, one that minimizes the degrees of freedom that describe the events producing the flow of stimulation. Analogous to the coordinative structures that combine articulators into the coordinative structures to produce gestural events, detection of higher-order invariants would automatically account for contextual variations such as speaking rate and style, allophonic variation due to phonetic context, speaker differences, and so on. Such invariants allow the perceiver to "hear through" lower-order variations that are irrelevant to phonetic coordinative structures in native speech. To illustrate, take the case of a man saying *Bob* normally vs. while clenching a pipe in his teeth. Bilabial closure for /b/ involves simultaneous jaw and lip narrowing movements, while the "ah" vowel involves jaw opening along with tongue body movement for pharyngeal narrowing. When the pipe is clenched, however, the jaws are held in a fixed, nearly-closed position. As a result, the speaker must accomplish the bilabial closure solely with the lips, and the vowel gesture solely with the tongue. The lower-order articulatory invariants of specific jaw, lip and tongue positions at specific times would thus differ between the two utterances, which together permit an attentive listener to hear whether the speaker's teeth are clenched. But the higher-order *phonological* invariant in both utterances is that bilabial closure occurs at both ends of the utterance, and a pharyngeal narrowing occurs between the two closures. Thus, the word *Bob* is perceived in both cases (i.e., the listener "hears through" the lower-order differences to detect the phonological structure). The higher-order description provides "reduced" information, relative to the lower-order one, by capturing fewer individual degrees of freedom.

The perception of non-native speech sounds by the native-language-educated attention of mature listeners would certainly be influenced by the perceiver's seeking of familiar higher-order invariants. In other words, the flip side of the efficiency of extracting native higher-order invariants may be an increase in difficulty of essentially "going back down a notch" to pick up the lower-order, and therefore more numerous, gestural details in unfamiliar non-native categories and contrasts which are irrelevant to critical distinctions among native gestural constellations (for further discussion of implications for second-language learning, see Best, in press b; cf. Flege, in press).

Although language-specific higher-order invariants are present in native speech, reflecting the coordinative structure among the distal articulatory events that produced it, most or all of these are initially beyond the perceptual reach of infants. They must still discover how the lower-order invariants of the simple articulatory components of gestures, which they are able to detect from early on, are harnessed into higher-order coordinative structures or gestural constellations by native speakers. Perceptual learning of the critical relational properties of higher-order structural and transformational invariants in native speech should thus entail a progressive reduction in the quantity of stimulus detail that must be detected, analogous to the reduction in directly-controlled degrees of freedom that results from the formation of coordinative structures in motor skill acquisition (or coordinated control of marionette limbs). This occurs because infants actively explore utterances to discover the optimal sets of gestural invariants that specify the native language structures that are interesting and useful to them. The latter, of course, continue to change as the infant develops, the discovery of lower-order invariants permitting the further discovery of higher-order ones.

By this ecological account, then, to learn to perceive the sound pattern of the native language, i.e., its phonological structure, is to discover the critical invariants specifying the various nested levels of gestural constellations in native speech. Learning to detect the crucial higher-order invariants means, of course, that there will be developmental change in the perception of native speech categories and contrasts. But given the presumed ability to detect lower-order articulatory invariants early on, developmental change in the perception of native patterns may be apparent mainly as increased efficiency in extraction of critical invariants. This increased efficiency may foster the infant's emerging ability to recognize words—sound-meaning relations—by the third quarter of the first year. That is, the infant should more easily and rapidly recognize the crucial gestural properties that define a given word irrespective of the irrelevant variation in its specific details when it occurs in different speech contexts, is produced by different speakers, etc. But perceptual learning of native gestural constellations also carries implications for developmental change in perception of non-native phonetic patterns during the same period. Developmental changes in perception of non-native sounds should be, and are, more dramatic because when the infant begins to discover language-specific invariants in native speech, he/she will pick them in native speech but will often be unable to find those familiar invariants in non-native utterances.

We turn now to the Perceptual Assimilation Model (PAM), which I developed to account for the developmentally-changing effect of experience with a particular language on the perception of non-native phonetic contrasts (Best, 1993, in press a, b; Best, McRoberts, & Sithole, 1988; Best & Strange, 1992). I began developing this model several years ago in an attempt to provide a coherent theoretical account for a number of observations in the literature on adult cross-language speech perception and on developmental changes in infant speech perception. Specifically, as indicated at the beginning of the chapter, adults often have difficulty discriminating non-native phonetic contrasts, while young infants have no such difficulty. Before the end of the first year, however, infants also begin to display difficulties discriminating non-native contrasts. However, no existing theoretical treatment offered a single, comprehensive explanation for 1) why, exactly, language-specific effects might occur in either adults or infants, 2) whether and why the effects might differ between adults and older infants, and 3) what the effects might suggest about the influence of phonological knowledge on perception. Certain complexities in reported adult findings would also have to be accounted for: discrimination levels appear to vary among different types of non-native contrasts, perception of non-native contrasts can be improved somewhat through perceptual training or through second language learning but this also depends on the type of contrast involved, and discrimination of non-native contrasts can be strongly affected by various task manipulations (the findings are reviewed and discussed in greater detail in Best, 1993, in press a, b).

Based on the considerations laid out in the preceding portion of this chapter, I used the ecological theory of perception as the foundation for developing a coherent theoretical account of the observations on cross-language speech perception in adults and infants. Thus, PAM is based on the ecologically-motivated assumption that efficient detection of native gestural patterns in speech may guide and constrain listeners' pickup of information in non-native phonetic categories and contrasts. This model is unique in several respects. First, it follows an ecological line of reasoning about perceptual learning rather than relying on innate linguistic abilities, information processing concepts, or cognitive development. Second, it attempts to provide a unified account for both adult cross-language perception findings and developmental changes in infancy. Third, it is the first to provide a detailed, coherent basis for predicting which non-native contrasts should be difficult to discriminate and which should be easy, and why. To the extent that PAM is compelling and is able to coherently account for the phenomena of cross-language speech perception in adults and infants, it obligates us to give serious consideration to the ecological approach.

We will turn next to an overview of how PAM accounts for the perception of non-native phonetic patterns by adults. For readers who are familiar with PAM, I should point out that there are several new features, by comparison with earlier versions of the model (i.e., Best, McRoberts, & Sithole, 1988; Best, in press a). Specifically, the relation between assimilation of non-native segments and discrimination of non-native contrasts has been clarified, additional discrimination types are now recognized and described, and the developmental aspects of the model are more fully delineated.

Perceptual assimilation model

The basic premise of the Perceptual Assimilation model (PAM) is that adults actively seek higher-order-invariants in speech which specify familiar gestural constellations, whether confronted with native or non-native utterances. Therefore, what they will perceive in non-native speech, at least initially when they have had little or no linguistic experience with the language involved, are the similarities and dissimilarities between the non-native gestural patterns and the familiar gestural constellations of their native language's phonological system (for more traditional accounts of the related phenomena of code-switching and loan-word phonology, see Elman, Diehl, & Buchwald, 1977; Silverman, 1992). For non-native phonetic patterns whose gestural organization is reasonably similar to the gestural invariant for one or more native phonetic categories, the adult listener is likely to detect native gestural invariants, and the non-native sound will be perceptually assimilated to the most similar native category(s). At the same time, however, listeners should also detect certain discrepancies between non-native phonetic patterns and native gestural constellations. After all, they are quite sensitive at detecting foreign accented utterances of their native language (Flege, 1984; Flege & Fletcher, 1992) and non-native dialect accents.

Note that these predictions are quite open to the possibility of individual differences among listeners regarding which invariants and discrepancies are detected, and how readily.[7] This is because non-native gestural constellations are not, of course, exactly the same as the native constellations but only resemble them more or less, i.e., they display similarity relations rather than identity relations. The resemblances are generally only partial; indeed, a given non-native gestural pattern may resemble more than one native constellation. Perception of the cross-language similarities would thus ride on selective attention, which is dependent on the listener's history of perceptual learning with the native language—for example, the particular invariants one learns could vary with the style and breadth of native utterances with which one has been engaged—as well as with other languages or other dialects of the native language (e.g., Chambers, 1992).

For consideration of the possible ways in which listeners may perceive non-native phonetic patterns, it is useful to conceptualize the ***native phonetic domain*** as the range of vocal tract sounds that are globally speechlike in their gestural properties, *vis a vis* the types of gestures and constellations employed in the native inventory of phonetic categories (for further development of this concept, see Best, in press b). Outside of this domain, in ***non-phonetic space***, are vocal tract-generated sounds such as coughs, chokes, laughs, whistles, razzes ("raspberries"), tongue clucking, squeals, etc. The latter three, and other non-speechlike vocalizations, occur in infant babbling and sound play. However, many infant vocalizations seem at least globally speechlike to adults, some being quite similar to native categories (as in /baba/ or /didi/) whereas others sound foreign, not falling clearly in any particular native categories (e.g., for and English speaker, the latter might include guttural sounds, tongue trills, etc.) (Oller, 1980; Oller & Lynch, 1992; Stark, 1980).

Analogously, there are three broad ways in which a non-native phonetic segment may be perceived with respect to the native phonetic domain (see Table 1). First, the perceiver may detect some resemblance to the gestural invariant of a native category (or perhaps more than one), in which case the non-native sound is perceptually assimilated to the native category, i.e., is categorizable. In cases of assimilation to a native category, the non-native segment may be virtually identical to the native gestural constellation, such that no cross-language discrepancy is perceived.

Table 1. *Perceptual assimilation of non-native phonetic segments.*

1. ***assimilated to a native phonetic category***
    a. identical to native gestural invariant:
        native sound
    b. reasonably similar to native invariant:
        acceptable exemplar of native category
    c. somewhat similar to native invariant, but noticeable discrepancies:
        deviant exemplar of native category
2. ***falls in unfamiliar region of native phonetic domain, outside any native categories:***
    unclassifiable speech sound
3. ***falls in non-phonetic space, beyond the boundaries of the native phonetic domain***
    nonspeech sound

Alternatively, the non-native segment may be somewhat discrepant but still sufficiently similar to be perceived as a good or acceptable exemplar of the native category. Or it may be even more obviously discrepant and thus be perceived as a poor exemplar of the category. Second, the non-native segment may be perceived as globally speechlike, but its gestural organization may not resemble any particular category in the native inventory very clearly. In this case, it will be perceived as speechlike but will not be assimilated to a specific native category. Rather, it will fall in an unfamiliar area of the phonetic domain and be an uncategorizable speech sound, as are the foreign-sounding elements in infant babbling. Third, the non-native segment may fall entirely outside the gestural range of the native phonetic domain and thus fail to be assimilated as speech, falling instead in non-phonetic space. These segments are non-assimilable as speech, and so will be perceived as nonspeech events, e.g., as nonspeech mouth sounds, snaps, clicks, etc.

However, the assimilation of individual non-native segments with respect to categories in the native inventory only touches the surface of the phonological component of the listener's language-specific grammar. Phonology encompasses the systematic functional relations among phonetic forms within a language, including distinctive segmental contrasts, allophonic alternations, phonotactic constraints, and other phonological processes (e.g., Jakobson & Halle, 1957; Silverman, 1992). From the ecological perspective on perceptual learning, the invariants that determine category membership differ qualitatively from the higher-order relational invariants which capture the critical differences that define the systematic relationships among categories. Thus, perceiving category membership can be more basic than recognizing critically distinctive relationships between categories. That is, one can recognize a particular instance of /b/ as an exemplar of the /b/ category because it has a complete bilabial closure and concurrent glottal vibration, without necessarily grasping that the critical difference from /d/ is constriction location.

For category membership, the perceiver may begin by extracting a set of lower-order properties of category members. But critical comparisons between categories depend on the abstraction of higher-order invariants that conjointly acknowledge the similarities that make comparison possible and capture the differences which crucially set the categories apart with respect to some purpose, such as a phonological contrast that serves to differentiate word meanings (J. Gibson, 1979). A critical contrast between events is characterized by *distinctive* features. Distinctive features do not merely list the lower-order properties of the individual classes, but rather they capture the relations between classes which remain invariant over contexts and non-identity-changing transformations, and which thereby define the uniqueness of each class with respect to the other (E. Gibson, 1963). The distinctive higher-order invariants that define phonetic contrasts indicate mere 'otherness' and cannot be heard independently of a speech segment, (E. Gibson & J. Gibson, 1972), e.g., location of constriction in the example above. Thus, they are more economical than category-defining properties, and optimize information pickup by an experience-attuned perceiver.

For these reasons, the influence of the systematic functional relations within the native phonology should be more readily apparent in perceptual comparisons between contrasting non-native categories than in a perceptual response to a single non-native category. As summarized in Table 2, PAM predicts that listeners will easily discriminate between non-native categories when they can detect in those sounds an invariant that specifies a critical difference, or phonological contrast, between gestural constellations in the native language (referred to as a Two-Category assimilation type, or *TC*). They should discriminate moderately well to very well between a non-native category for which they detect strong similarity to a given native gestural constellation and another non-native category for which they detect less similarity (or greater discrepancy) to the same native category (Category Goodness difference, or *CG* assimilation type), or versus one for which they cannot detect clear similarity to any single native constellation (Uncategorized vs. Categorized assimilation type, or *UC*). When the non-native categories both bear only a global resemblance to the gestural constellations of (native) speech but do not assimilate clearly into any particular native phonetic category(s), they will be both assimilated as uncategorizable speech sound (both Uncategorizable, or UU), and will be moderately to fairly difficult to discriminate, depending on they bear any remote similarity to any native category(s) and the extent to which any such similarities overlap between the two non-native sounds. Discrimination should also be very difficult when both members of the non-native contrast are perceived to fit *within* a gestural constellation for a single native category equally well (Single Category assimilation type, or *SC*). The SC case and the CG case actually fall at different points along a single dimension, in that both involve non-native contrasts whose members are assimilated to a single native category. Thus, to the extent that prototype effects in perception of phonetic categories (i.e., asymmetries in discrimination around good vs. poor exemplars of a category—e.g., Grieser & Kuhl, 1989; see description in next section) are operative in speech perception, they should combine with the SC and CG assimilation patterns to predict better SC discrimination when both non-native categories are assimilated as poor (non-prototypical) rather than as good exemplars of the native category, and to predict CG discrimination asymmetries that reflect greater category generalization (poorer discrimination) around prototypical exemplars than around non-protoypical exemplars of the native category.Discrimination should be moderately to very good, comparable to the CG assimilation type, if both non-native gestural patterns are perceived to fall outside the native phonetic domain altogether, in non-phonetic space (Non-Assimilated type, or *NA*).

Table 2. *Assimilation effects on discrimination of non-native contrasts.*

Contrast Assimilation Type	Discrimination Effect
Two-Category (TC)	excellent discrimination each non-native sound is assimilated to a different native category
Category-Goodness Difference (CG)	moderate to very good discrimination both non-native sounds assimilated to the same native category, but they differ in discrepancy from native "ideal" (e.g., one is acceptable and the other is deviant) can vary in degree of difference as members of native category
Single-Category (SC)	poor discrimination both non-native sounds assimilated to the same native category, but are equal in fit to the native "ideal" better discrimination for pairs with poor fit (equally poor) to native category than pairs with good fit (equally good)
Both Uncategorizable (UU)	poor to moderate discrimination both non-native sounds fall within unfamiliar phonetic space can vary in their discriminability as uncategorizable speech sounds
Uncategorized vs. Categorized (UC)	very good discrimination one non-native sound assimilated to a native category, the other falls in unfamiliar phonetic space, outside native categories
Non-Assimilable (NA)	good to very good discrimination both non-native categories fall outside of speech domain and are heard as non-speech sounds can vary in their discriminability as nonspeech sounds

The earlier comparisons of gestural scores for English and non-English phonetic categories illustrate some of these cross-language gestural similarities and dissimilarities. In the Hindi [d̪ɑ]-[ɖɑ] example (Figure 3), the dental versus retroflex constriction locations do not distinguish English stop consonants; in fact, they occur as phonologically equivalent (i.e., non-distinctive) allophonic variants of (alveolar) /d/. As for the Zulu [k^h]-[k'] example (Figure 4), a distinctive property of the voiceless velar stop in English [k^h] is a glottal opening gesture coordinated with closure (as in Zulu [k^h]). This critical gesture is lacking from Zulu [k'], which instead has a glottal closure and is therefore notably discrepant from [k^h]. The Zulu voiced-voiceless lateral fricatives differ by essentially the same glottal voicing distinction (open glottis versus critically closed glottis) found in similar English fricative contrasts (e.g., /s/-/z/, "sh"-"zh"). Lastly, the dual alveolar+velar closures and the suction release gesture for Zulu alveolar versus lateral clicks are globally unlike anything in English phonology, and resemble nonspeech events such as cork popping and finger-snapping rather than being even generically speechlike for most English listeners.

PAM thus predicts that adults' attunement for detecting the articulatory gestural invariants that specify familiar phonetic categories of the native language will foster detection of both similarities and dissimilarities between non-native segments and the native inventory. Even more importantly for questions about perceptual influences of the native phonological system, discrimination of non-native *contrasts* is predicted to depend on the listener's abstraction of higher-order invariants that specify distinctive oppositions in the native phonology, as well as on their detection of discrepancies between the native contrasts and gestural properties of contrasting non-native segments. But what of young infants, who are not yet perceptually attuned to native phonetic categories, and especially to the native phonological system? When and how do infants begin to extract the gestural invariants of native categories and the higher-order invariants of critical distinctions found in native contrasts? And how does this early perceptual learning of the phonetic categories and relationships of the native language begin to affect perception of non-native phonetic forms?

To provide a basis for discussing these issues, we will begin with a brief review of empirical findings on developmental changes in infants' perception of native and non-native phonetic contrasts. Following that, we can outline a perceptual learning account of development that appears to accommodate those facts. That outline will provide the background for studies I have conducted with students and colleagues to test several predictions of PAM for perception of varying non-native phonetic contrasts by adults and infants.

Developmental changes in infant perception of phonetic contrasts

Young infants, up to about 4 months of age, have had relatively limited experience hearing the native language. Even the language experience they have had generally focuses attention more on prosodic patterns than on minimal segmental contrasts. The infant-directed speech that is typically addressed to them is characterized by exaggerated pitch contours and durational properties, relative to adult prosody, in most cultures (Fernald et al., 1990; Fernald & Mazzie, 1991; Fernald & Simon, 1984; Grieser & Kuhl, 1988; cf. Bernstein Ratner & Pye, 1984). Moreover, infants from birth to at least 4 months of age prefer listening to infant-directed speech more than to adult-directed speech (Cooper & Aslin, 1990; Fernald, 1984, 1985; Fernald & Kuhl, 1987; Werker & McLeod, 1990). In contrast with its prosodic properties, infant-directed speech is not marked by exaggeration or emphasis of segmental distinctions (Bernstein Ratner, 1984, 1986; Bernstein Ratner & Luberoff, 1984; Malsheen, 1980). Even so, many findings indicate that young infants do discriminate a broad range of consonant and vowel contrasts in nonsense syllables, regardless of whether or not the contrasts occur in their language environment (e.g., Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Jusczyk & Thompson, 1978; Jusczyk, Copan, & Thompson, 1978; for comprehensive reviews, see e.g., Aslin, 1987; Aslin, Pisoni, & Jusczyk, 1983; Best, 1984; Kuhl, 1987; Jusczyk, 1994). Evidence for developmental decline in discrimination of certain non-native contrasts will be discussed in depth in a subsequent section.

A few phonetic differences have been suggested to pose difficulties for young infants, *viz*, certain native fricative voicing contrasts (e.g., English /s/-/z/: Eilers, 1977; Eilers & Minifie, 1975) and fricative place contrasts (e.g., /f/-"th" [*think*:] Eilers, Wilson, & Moore, 1977). However, more recent work by those researchers, as well as by others, has shown that infants do discriminate those same contrasts (Eilers, Gavin, & Oller, 1982; Holmberg, Morgan, & Kuhl, 1977; Levitt, Jusczyk, Murray, & Carden, 1988). Moreover, infants discriminate other fricative place of articulation contrasts, both native (e.g., /s/-"sh": Eilers, & Minifie, 1975; Eilers, Wilson, & Moore, 1977; Kuhl, 1980) and non-native (e.g., Czech retroflex vs. palatal voiced fricatives: Trehub, 1976; Eilers et al., 1982). The balance of that evidence indicates that young infants can discriminate native and non-native fricative contrasts.

In addition to the basic discrimination findings, infants under 4 months show other revealing perceptual patterns. When familiarized with a set of syllables that share either a common vowel and different consonants, or the converse, 2-month olds and newborns can detect the addition of new syllable that differs in either consonant or vowel or both (e.g., Bertoncini, Bijeljac-Babic, Jusczyk, Kennedy, & Mehler, 1988; Jusczyk & Derrah, 1987), although newborns are more affected by attentional manipulations (Jusczyk, Bertoncini, Bijeljac-Babic, Kennedy, & Mehler, 1990). This pattern suggests that young infants perceive the syllables holistically rather than as a combination of discrete segments. Infants between 2-4 months can also discriminate 3-5 syllable utterances whose medial syllables differ, but apparently only if the contrasted elements are highlighted by the exaggerated prosodic contours of infant-directed speech, or differ on more than one articulatory feature (e.g., /r/-/k/) (Goodsitt, Morse, Ver Hoeve, & Cowan, 1984; Fernald & Kuhl, 1982, cited in Karzon, 1985; Karzon, 1985; see review by Jusczyk, 1993).

Vowel prototype, or "magnet," effects may also be found quite early. The magnet effect refers to a perceptual pattern in which listeners show preferences for and greater generalization (poorer discrimination) around good rather than poor exemplars of a vowel category (as per adult goodness ratings) (Grieser & Kuhl, 1989). These perceptual asymmetries around good versus poor tokens indicate that perception of vowel categories is not absolute, but rather shows systematic within-category differentiation, an effect which occurs only in humans and not in monkeys (Kuhl, 1991). The discrimination asymmetry for good vs. poor tokens has been found in human newborns with both native and non-native vowels (Walton & Socotch, 1993). By 6 months of age, infants still show the effect for a native vowel (Grieser & Kuhl, 1989) but not for a non-native one (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Polka & Werker, in press) (the latter findings will be discussed in more detail later).

In addition, young infants are able to perceive, for at least some consonants and vowels, an underlying phonetic category identity throughout the variations introduced by different pitch contours, different speakers and different adjacent segments. Detection of such a phonetic equivalence class would appear as perceptual constancy across such variations in a phonetic category, within the familiarization or background stimuli and within the test stimuli. Perceptual constancy was shown in 1-4 month olds for discrimination of a vowel contrast presented with pitch contour variations (Kuhl, 1979; Kuhl & Miller, 1982). Similar perceptual constancy in discrimination of a consonant contrast across speaker variations has been found in 2 month olds (Jusczyk, Pisoni, & Mullennix, 1992), but only if there is no delay between the familiarization and testing phases. Similar memorial effects have been found in adults (Martin, Mullennix, Pisoni, & Summers, 1989). Perceptual constancy across varying phonetic contexts (e.g., /p/ across /pi/, /pɑ/, /pu/; nasalization across /nɑ/, /mɑ/, /ŋɑ/) has been found for both vowels and consonants by 4-6 months of age (e.g., Fodor, Garrett, & Brill, 1975; Hillenbrand, 1983, 1984; Kuhl, 1979, 1980, 1983). Thus far, only native phonetic categories have been tested with infants.

The findings summarized thus far have demonstrated little evidence of developmental changes in basic aspects of infant speech perception for native segmental contrasts, save for some signs of increased susceptibility to attentional manipulations or memorial disruptions in the first two months (Bertoncini et al., 1988; Jusczyk, Pisoni, & Mullinnex, 1992). However, in final quarter-year, there are some clearer indications that perception of native segmental patterns is beginning to be influenced by experience with the language. As discussed earlier, languages differ in both the inventories of consonants and vowels they employ, and also in their phonotactic rules regarding permissible sequencing of those elements. When 9 month olds are permitted to choose between listening to two series of unfamiliar words with English vs. Dtuch segments and phonotactics, infants from each language preferred listening to the list representing their native language. Younger infants showed no preference between these prosodically-similar languages. Although English-learning infants did show a native preference when presented with English vs. prosodically-different Norwegian, that effect was solely attributable to prosody rather than segmental and phonotactic constraints (Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993). The experiential effect on 9 month olds' preference for segmental patterns is strengthened by recent findings that Dutch infants this age prefer phonotactically permissible vs. phonotactically impermissible sequences of Dutch segments (Friederici & Wessels, in press), and that American infants prefer frequently-occurring vs. infrequently-occurring English phonotactic patterns (Jusczyk, Charles-Luce, & Luce, submitted—see Jusczyk, in press).

Infants' discovery of relations between sound patterns and meaning also begins around last quarter of first year, with the beginnings of word comprehension. Infants usually begin producing single words a few months later, at around 12-13 months on average, followed by the emergence of syntactic abilities with their first simple word combinations at around 18 months. A phonetic contrast that young infants discriminated in simple discrimination tests, prior to the emergence of word comprehension, may later be missed altogether as a minimal phonological contrast by the one year old whose comprehension vocabulary still lacks minimal word pairs (e.g., the /d/-/b/ contrast when it appears in *dog* vs. *bog*). This follows from the claim of child phonologists that the earliest linguistic units in the single-word period of child speech are more global than the segment (e.g., Ferguson, 1986; Ferguson & Farwell, 1965; Macken, 1992; Macken & Ferguson, 1983; McCune, 1992; McCune & Vihman, 1987; Menn, 1986; Menn & Matthei, 1992; Vihman, 1992), and that segments are gradually differentiated in both production and perception from these early, more global units (e.g., Goodell & Studdert-Kennedy, 1990; Lindblom, MacNeilage, & Studdert-Kennedy, 1983; Nittrouer, Studdert-Kennedy, & McGowan, 1989; Studdert-Kennedy, 1986, 1991) due to the pressure exerted by vocabulary expansion on the organization of the lexicon (Lindblom, 1992; Studdert-Kennedy, 1987, 1991). Discrimination of minimal contrasts in meaningful word contexts appears to emerge around 18-19 months of age (Werker & Baldwin, 1991; see Werker & Pegg, 1992). Similar temporary dips in phonetic ability have also been noted in early word productions, where they are taken as evidence of progress in the development and systematization of phonological knowledge (e.g., Macken, 1992; Menn & Matthei, 1992).

In the next section, I will outline the perceptual learning framework for development of speech perception in infancy (and somewhat beyond). The suggested path of learning is informed, in part, by the findings summarized above, in addition to the general principles of the ecological approach to perception. It provides the backdrop for considering the research findings on adults' and infants' perception of non-native phonetic contrasts, particularly a series of studies motivated by PAM, which will be described in the subsequent section.

Perceptual learning and infant speech perception

The basic assumption of the ecological account of perceptual learning offered here is that the type of gestural information the child perceives in speech will change developmentally with increasing attunement to the ambient language. The infant will become better able with experience to detect both finer structure and more encompassing structures in native utterances. Following Eleanor Gibson's (1991) arguments about perceptual learning in general, the detection of gestural patterns in speech should become increasingly specific to the phonological categories and contrasts of the native language, there should be an increasing optimization of attention to them, and pickup of gestural information should become increasingly economical, that is, focus should shift away from irrelevant properties and sharpen for critically distinctive ones. The distinguishing features detected for discrimination should shift developmentally, showing progressive improvement in finding the critical features and in abstracting higher-order invariants, both of which reduce the number of comparisons required for discrimination (E. Gibson, 1969; 1971). These are exactly the advantages afforded to an experienced listener by the phonology of the native language. Because the language-specific phonological system reduces lower-order phonetic detail to just those distinctive features that are crucial for grammatical purposes (e.g., Archangeli, 1988) and organizes that information into superordinate structures, it allows a sensitized perceiver to take in more information within a given time frame and to minimize uncertainty about the important linguistic units. As experience with the native language optimizes and economizes information pickup, therefore, the infant begins to discover the phonological principles of that language.

This learning will, in turn, be reflected in developmental change in the infant's perception of nonnative categories and contrasts. Progress in perceptual learning about the native language should result in, and be illuminated by, developmental changes in perception of non-native speech. The suggested pattern of perceptual learning about native phonological structure, and its expected effects on infant's perception of non-native categories and contrasts, is summarized in Table 3.

During about the first quarter-year of life, very young infants should have attained minimal perceptual learning of the higher-order invariants for native segmental contrasts, at best. Their experience with the native language is relatively limited, and the speech typically addressed to them generally focuses attention more on prosodic patterns than on minimal segmental contrasts. The view posited here is that infants initially detect simple differences in low-order articulatory invariants, such as the velar versus alveolar closure location for /g/-/d/, the presence versus absence of a glottal opening gesture for /p/-/b/, or the high versus slightly lower tongue position near the front of the vocal tract for "ee"-"ih". This ability should extend to simple gestural differences in both native and non-native phonetic contrasts.

Table 3. *Perception of native and non-native contrasts in infancy and early childhood.*

developmental phase	**information detected**	**native phonetic categories**	**Non-native phonetic categories**
1st quarter-year (0-3 months)	simple articulatory gestures *(language universal)*	discriminates any vowel & consonant difference	same as for native speech
	good vs. poor exemplars of simple gestures *(language universal)*	prototype effects for vowels (and consonants?)	same as for native speech
	invariants of simple gestures under speaker & intonation variations *(language universal)*	perceptual constancy for vowels and consonants	same as for native speech
2nd quarter-year (3-6 months)	continues as above (language universal)	continues as above	same as for native speech
	invariants of simple gestures under phonetic context variations (language-specific?)	perceptual constancy for native categories	may fail with non-native categories
3rd quarter-year (6-9 months)	simple relational invariants for vowels *(language-specific)*	discriminates native vowel differences	fails to discriminate non-native vowels that differ from native relational invariants
	good vs. poor vowels re: relational invariants	prototype effects for native vowel categories	lacks prototype effect for non-native relational invariants
4th quarter-year (9-12 months)	simple invariants for native gestural constellations *(language-specific)*	discriminates native vowel and consonant categories	discriminates if able to detect different native invariants, or good vs. poor native invariant or if no speechlike gestures at all
		prefers listening to common native syllable patterns more than non-native or uncommon native patterns	fails if detects a native invariant but not a goodness difference or if detects speechlike gestures but not any native invariants
extending to 2nd year (9-17 months)	simple invariants for sound-meaning association	learns to recognize simple native words and meanings re: global gestural patterns	may have difficulty learning meaning associated with non-native global patterns
18 months	higher-order relational invariants for minimal contrast word pairs	detects native phonological contrasts	perception of non-native phonological contrasts depends on similarity to native contrast invariant
2 - 5 years	higher-order relational invariants among some allophones higher-order invariants specifying morphological alternations, etc.	tendency toward perceptual equivalence among allophones of a category	no difference in response to non-native allophones vs. non-native phonol. contrasts

Given the assumption that they detect simple differences in low-order articulatory invariants, it should not be surprising that infants in the first quarter-year can pick up simple gestural commonalities within phonetic categories even in the face of certain category-irrelevant variations. That is, they show perceptual constancy for simple phonetic equivalence classes across non-identity-changing transformations. Because lower-order articulatory invariants of phonetic categories are not greatly affected by speaker (within a single dialect) and intonation variations, but may be affected by phonetic context variations due to coarticulation of consonants and vowels, perceptual constancy across speakers and intonation patterns may be evident earlier in development than perceptual constancy across different phonetic contexts. Thus far, the phonetic constancies demonstrated in the first quarter-year (Jusczyk, Pisoni, & Mullennix, 1992; Kuhl, 1979; Kuhl & Miller, 1982) have involved only speaker and intonation variations. Only the studies with infants in the second quarter-year (Fodor, Garrett, & Brill, 1975; Hillenbrand, 1983, 1984; Kuhl, 1980, 1983) have involved phonetic variations. In addition, given the slower, longer-lasting, more global tongue gestures associated with vowels as opposed to the more rapid and localized constriction gestures associated with consonants, perceptual constancy may appear earlier for vowels, or may simply be more easily obtained and more robust to attentional manipulations, than constancy for consonants. Again, studies of very young infants (Kuhl, 1979; Kuhl & Miller, 1982) have tended to test only vowel constancy, whereas studies with infants in the second quarter-year (Fodor et al., 1975; Hillenbrand, 1983, 1984; Kuhl, 1980, 1983) have tested for consonant constancy. The possibility of a vowel vs. consonant difference also seems compatible with the findings of Bertoncini et al. (1988) and Jusczyk et al. (1992) (cf. Jusczyk et al., 1990). However, further investigation is needed to evaluate both possibilities of early developmental changes in perceptual constancy.

Regardless of these possible stimulus parameter effects on perceptual of phonetic equivalence classes, very young infants should show constancy equally for native and non-native phonetic categories. To the extent that phonetic categories and contextual effects differ among languages, infants should become attuned to native language patterns and we should expect to see some language-specific effects emerge later, probably around the second half-year. Thus far, however, no studies have examined phonetic perceptual constancy re: variations of speaker, intonation, or phonetic context in infants of any age.

The assumption that very young infants detect simple gestural properties of phonetic categories also admits the likelihood that they should also show so-called perceptual magnet effects within the first quarter-year, at least for vowels. This is based on the reasoning that prototypes and non-prototypes differ in how well they convey the important gestural properties of a vowel category. This, in turn, would affect how easily perceivers could detect the gestural pattern of the category in the differing stimulus tokens. The notion that there is an articulatory basis for good versus poor vowels is consistent with the quantal theory of speech. The quantal theory demonstrates that certain vowel types are very stable, in that small changes in their articulatory constriction location produce minimal changes in the acoustic pattern of the vowel, whereas other constriction locations are unstable acoustically. Languages tend to avoid the latter locations for possible vowels (Stevens, 1972, 1989). Infants in the first quarter-year would be expected to show magnet effects for both native and non-native vowels, a prediction that is consistent with one recent report (Walton & Socotch, 1992).

Young infants in the first quarter-year should not yet recognize the more complex coordination or phasing required for specific native gestural constellations, e.g., syllable-initial /l/ in English has an uvular narrowing gesture which follows the tongue tip closure gesture for /l/, rather than being synchronous with it as in word-final English /l/ and in the Russian "hard" /l/, or absent as in the Russian "soft" /l/. Only as infants become attuned to detecting invariants for familiar gestural constellations in native speech should they begin to show effects of native language experience on their perception of non-native contrasts. This sort of native attunement would not be expected until at least the second quarter-year (perhaps in perceptual constancy across phonetic context variation), or more likely the following quarter-year.

By the third quarter-year (second half-year), infants should progress to discovering and attending to more economical higher-order relational invariants found in the native phonology, such as the ratio of the two portions of the vocal tract that fall on either side of the tongue constriction location for a given native vowel. These discoveries are assumed to proceed systematically from less to more encompassing and more economical invariants. Thus, the first sorts of native relational invariants infants are likely to discover are relatively simple ones such as the ratio between the length of the vocal tract that lies before versus behind the high front tongue constriction for the vowel /i/ ("ee"). Once they detect such invariants, they should begin to show language-specific influences on perception of non-native vowel contrasts and prototypes. These older infants' abilities to discriminate non-native vowels and to perceive non-native vowel prototypes will depend on whether they can detect in those stimuli the relational invariants that they can now detect in native vowels, i.e., in whether they "assimilate" the non-native vowels to native categories. If so, performance will further depend on whether the infant assimilates the non-native vowels as good exemplars of native category, and whether two contrasting non-native vowels are assimilated to the same native category or to different categories. However, we should not expect infants' assimilations to match those of adults completely because infants' detection of native vowel invariants is surely not as well-tuned as that of adults, and the invariants they detect may be somewhat lower-order than those of adults.

With further experience, by the last quarter of the first year, infants should also begin to recognize the higher-order invariants that specify native gestural constellations for consonants, as well as the broader phonotactic patterns of native syllables. For example, they should begin to recognize the higher-order relational invariants that specify consonantal gestural constellations in the native language, such as the precise phasing between the bilabial closure and the glottal opening gestures for English /p/ (as opposed to the different phasing for French /p/). At this point, infants' listening preferences and discrimination abilities will reflect language-specific influences on perception of non-native consonants and syllable types (re: phonotactic rules regarding how consonants and vowels may be sequenced to form syllables). Older infants' perception of these sorts of non-native gestural constellations will also depend on whether and how those patterns provide the higher-order gestural invariants they have learned to detect in native consonants and syllable types. Again, these older infants' assimilations of non-native constellations to native categories is still not expected to match adults' assimilation patterns, which derive from a much more sophisticated level of perceptual learning that incorporates minimal phonological contrast and other even more complex relations among segments in the native phonology.

At this point in development, however, infants would not necessarily perceive allophones of a given phoneme as related variants of a single segment, such as the allophonic relationship among stressed syllable-initial voiceless aspirated /p/ versus unreleased final /p/ versus voiceless unaspirated /p/ after /s/. Instead, they may detect differences among allophones simply as gestural characteristics of differing native syllable patterns. This is because they presumably would *not* yet have discovered the even higher-order invariants that relate allophones to common underlying phonological categories. Such abstract commonalities draw on grammatical relations among lexical items (e.g., different morphological forms of a stem word—see further discussion below), which are still beyond young infants' grasp.

Sound-meaning associations, which relate the higher-order gestural constellation of the spoken word to the confluence of contextual signs of its meaning, emerge in comprehension during the final quarter-year. Some ecological, perceptual learning accounts of this important discovery have been offered in the literature. For example, parents often repeat a key word several times to their infant under diverse spoken transformations, such as variations in prosody and sentence frame, while they concurrently engage the named object (noun) in different event transformations such as holding it out or wiggling it back and forth, or while they produce variations on the named action (verb)(Dent, 1990; Dent & Rader, 1979; Goldring Zukow, 1991; Zukow & Schmidt, 1988). The articulatory gestural component infants extract for such sound-meaning complexes is expected to be less differentiated phonetically than other gestural patterns the same infant might detect in the absence of a sound-meaning relation, because the added dimension of semantic or contextual information for words must be reconciled with the limitations of the infant's perceptual span and the need for economization of information pickup. For this reason, children's early words, in both production and perception, should be differentiated by rather holistic gestural properties and not by the finer grain of minimal contrasts (see Best, in press a). Minimal contrasts that they discriminated prior to the emergence of meaning are likely to be missed now in sound-meaning complexes. Infants at this point have still not discovered minimal phonological opposition. Discovery of phonological oppositions *per se* requires detection of finer-grained distinctions between the gestural constellations of minimally contrastive, meaningful lexical items. The ability to perceive phonological contrasts as such may not be apparent until the upper edge of the infancy period. Recall that minimal contrasts are part of the phonological component of a language-specific grammar. The perception of minimal contrast in the native language, a minimum requirement of a segmental phonology, should be associated with the so-called spurt in children's productive vocabulary (>50 words), which also predicts the emergence of syntax and morphology (e.g., Macken, 1992). At that point, the comprehension vocabulary, if not also the production vocabulary, should be large enough to include minimally contrastive word pairs such as *bed-bad* or *peas-keys*. To perceive a phonological contrast a relational invariant must be extracted, the critical segmental distinction that marks a difference in meaning between a minimal pair of words. This characterization is consistent with the earlier-summarized finding that older infants begin to detect minimal contrasts in meaningful words around 18-19 months (Werker & Baldwin, 1991; see Werker & Pegg, 1992).

Discovery of the still higher-order invariants corresponding to numerous other aspects of phonological structure await still more experience with the native language, some probably requiring years. For example, perceptual learning of allophonic relations should depend in part on hearing the same word produced by different speakers and with varying speech styles (e.g., casual, formal, and careful speech), as well as on hearing how morphological operations on words affect the phonetic form of the base word. To illustrate, in American English casual speech /t/ and /d/ have a number of context-conditioned allophonic variants: unreleased stops in final position (e.g., *sit, dad, mad*); rapid tongue taps (flaps) as onsets of non-initial unstressed syllables (*sitting, daddy, kitty*); glottal stops or nasal-released stops preceding unstressed syllabic /n/ (*kitten* versus *hidden*, respectively). Word pairs that young children are likely to hear could provide them with evidence of some of these phonological relations, as in the unreleased /t/ versus flap in *sit-sitting*, the unreleased final /d/ versus medial flap in *dad-daddy*, the flap versus glottal stop in *kitty-kitten*, and the unreleased /d/ versus nasal-release in *hid-hidden*. In these cases morphological transformations of meaningful, known words provides a crucial link among the diverse allophones. Adults may also help clarify some allophonic relations if they "correct" their normal conversational speech patterns by repeating words in careful, precise speech to young children. To illustrate, although they pronounce *kitty* conversationally with a medial flap, they may at times pronounce it carefully for the child (as when correcting the child's spelling errors), with the medial /t/ as a voiceless alveolar stop (see Bernstein Ratner, 1993). An underlying gestural commonality among the diverse allophones of medial /t/-/d/ is apparent in children's productions by 20-22 months (Best, Goodell, & Wilkenfeld, in preparation; Best, in press a). More abstract phonological relations among allophones may also be highlighted later by learning to read and spell, as is the case for the flapped allophones of /t/ and /d/ (see Treiman, Cassar, & Zukowski, submitted).

Similarly, children may learn about even more abstract phonological relations through frequently used morphological operations. For example, the English voiced-voiceless alternations between /s/-/z/ in noun pluralization (e.g., *cats* versus *dogs*) and between /t/-/d/ in the past tense forms of regular verbs (e.g., *walked* versus *climbed*) covary with the voicing of the preceding segment. Morphological development during the preschool years (Berko, 1958) should aid children's discovery of related phonological alternations (see also Gerken, Landau, & Remez, 1990; Gerken & McIntosh, 1993). Other structural properties of the native phonological system that may take even longer for the child to fully apprehend in speech include some aspects of linguistic stress and intonation, for which perceptual learning may extend to as late as 7-10 years (Cruttenden, 1974).

In the next section, I will review recent data from my own and others' laboratories that pertains to the preceding account of perceptual learning about the native phonology, and its influence on perception of unfamiliar non-native phonetic contrasts. The findings will be discussed within the framework of the Perceptual Assimilation Model (PAM), though it should be noted that the work of other researchers was generally not motivated by PAM. Although much of the research has involved consonant contrasts, some more recent work focuses on vowel contrasts; these areas will be described in separate subsections below. Because PAM's assimilation and discrimination predictions were developed to account for *mature* listeners' perceptions of non-native phonetic contrasts, adult findings will be described first within each area.

Experimental evidence on PAM and development of perceptual learning

Consonant contrasts. PAM predicts that adults' ability to discriminate different non-native contrasts will vary depending on how they assimilate the non-native phonetic categories *vis a vis* the phonological inventory of their native language.[8] The assimilation predictions presented here and elsewhere (see Best, 1993, in press a; Best et al., 1988; Best & Strange, 1992) refer specifically to adults' initial perception of unfamiliar contrasts from languages with which they have had little or no linguistic experience. However, the model could be extended, via the principles of perceptual learning outlined here, to account for changes in perception that can occur as adults learn a second language (see Best, in press b; for an alternative view, see Flege, in press). To review PAM predictions briefly (Tables 1 and 2), adults are expected to show excellent discrimination for non-native contrasts that are assimilated to two different native categories (TC assimilation type). They should show good to very good discrimination for those that are not assimilated into native phonetic space (i.e., are heard as nonspeech: NA type), or for those assimilated with differing degrees of goodness into a single native category (CG type), or for those in which one pair member is assimilated to a native category but the other is uncategorizable (UC type). Moderate to poor discrimination is expected for non-native contrasts that fall within unfamiliar phonetic space (i.e., are both heard as uncategorizable speech sounds: UU type), and poor discrimination for those assimilated as equally good exemplars of a single native category (SC type).

Earlier reports of poor discrimination of non-native consonants by adults have tended to use contrasts that were most likely assimilated as SC types or perhaps as UU types. Discrimination levels for such contrasts should indeed have been low according to the Perceptual Assimilation Model. For example, speakers of Japanese and Korean who are relatively inexperienced with spoken English have great difficulty discriminating and differentially labeling English /r/-/l/ (e.g., Gillette, 1980; Goto, 1971; Miyawaki et al., 1975; Mochizuki, 1981; Sheldon & Strange, 1982; Yamada & Tohkura, 1991). Their languages do not have an /l/ category and their /r/ is not a liquid approximant as in American English, but rather a flap more like the medial /d/ in *daddy* (Bloch, 1950; Price, 1981; Vance, 1987). Thus, PAM would expect monolingual Japanese to assimilate both English /r/ and /l/, maybe as poor examplars of their flapped /r/, but more likely as poor exemplars of their approximant /w/ or as uncategorizable speech sounds. The sounds should be rather poorly discriminated by Japanese in any of these cases, although perhaps slightly above chance.

In a study conducted before the development of PAM, Kristine MacKain, Winifred Strange and I compared American and Japanese listeners' labeling and discrimination of /l/-/r/ in a computer-synthesized continuum ranging from English *rock* to *lock* in acoustically-equal steps (MacKain, Best, & Strange, 1981). As expected, the American listeners strongly displayed the phenomenon of categorical perception. That is, they labeled the items at one end of the continuum very consistently as /l/ and the items at the other end as /r/, with a steep category boundary. Correspondingly, their discrimination between items that were 3 steps apart along the continuum was poor for within-category comparisons but very good for between-category comparisons, with a dramatic peak in discrimination performance at the position of the category boundary found in labeling. Japanese who had had little English conversational experience, on the other hand, showed nearly flat labeling and discrimination functions, with no category boundary effect and poor discrimination overall. Interestingly, however, a subgroup of Japanese subjects who had had some period of intensive conversational training and/or practice in English showed labeling and discrimination functions similar to the Americans', although not quite as high. Thus, the results are compatible with PAM, and in addition suggest that perceptual of non-native contrasts can be improved by intensive conversational experience with the language involved (see also Flege, 1989, 1991a; other training approaches may also improve discrimination: e.g., Jamieson & Morosan, 1986; Logan, Lively, & Pisoni, 1991; Pisoni, Aslin, Perey, & Hennessey, 1982; Strange & Dittmann, 1984).

Monolingual English-speaking listeners have also, of course, shown poor discrimination for a number of non-native contrasts, each of which is most likely to show SC assimilation patterns. For example, Thai voiced vs. voiceless unaspirated utterance-initial stops are both good exemplars of English voiced stops, and are difficult for English listeners to discriminate (Lisker & Abramson, 1970). Hindi voiceless unaspirated dental vs. retroflex stops, which are likely to be heard as /d/, are quite difficult for English listeners to discriminate, as are Nthlakampx (Thompson: Interior Salish) velar vs. uvular ejective stops /k'/-/q'/, which are likely to be heard as "odd" exemplars of English /k/ (or sometimes as other English sounds) (Polka, 1991; Werker, Gilbert, Humphrey, & Tees, 1981; Werker & Lalonde, 1988; Werker & Tees, 1984a). Likewise, the Czech retroflex vs. palatal voiced fricatives are poorly discriminated by English listeners (Eilers, Gavin, & Oller, 1982; Trehub, 1976), who are likely to hear them both as "zh".

Also relevant to the perceptual learning approach more generally are several studies showing that reducing the memory demands of the discrimination task or "stripping away" all acoustic details other than the crucial difference between the contrasting non-native categories result in increased discrimination of SC type contrasts (e.g., Carney, Widin, & Viemeister, 1977; Miyawaki et al., 1975; Pruitt, Strange, Polka, & Aguilar, 1990; Werker & Logan, 1985; Werker & Tees, 1984a). Both experimental manipulations reduce the array of information within which the listener must detect the critical differences. With the acoustic manipulation in particular, in reducing or eliminating the irrelevant and redundant stimulus properties, the experimenter both picks out the distinctive features for the listener and simultaneously attenuates the speechlike properties of the stimuli, i.e., moves them toward NA assimilation types.

A more comprehensive examination of the Perceptual Assimilation model, however, requires the comparison of discrimination levels across differing non-native assimilation types, and direct assessment of the listeners' assimilations of the non-native sounds re: their native categories. The first study on this point investigated perception of several click consonant contrasts from Zulu, a southern African Bantu language, by American English adults who were completely inexperienced with any click languages (Best, McRoberts, & Sithole, 1988). Clicks should not be assimilable as speech sounds within English phonetic space because their manner and place of articulation are different from anything in the English inventory of gestural constellations. That is, the click contrasts should produce an NA assimilation pattern for most English listeners, and should be relatively easily discriminated as nonspeech sounds. Subjects were tested with multiple natural tokens on discrimination of all minimal-feature pairings from the three by three matrix of Zulu click voicing categories (voiceless, short-lag voiceless, voiceless aspirated) and places of articulation (alveolar, lateral, palatal), which yielded 18 minimal contrasts. According to posttest questionnaires, the listeners assimilated all clicks as various nonspeech sounds (e.g., "a cork popping," "tongue clucks," "finger snaps"), except for one subject who heard some clicks as being similar to English /k/. Performance on an AXB discrimination test was quite good, ranging from 80% correct (chance = 50%) for the most difficult contrast, the alveolar vs. lateral voiceless unaspirated pair, to 85-95% correct for the others. Thus, the PAM prediction of good to very good discrimination for non-native NA contrasts was met, and performance differs substantially from that reported above for non-native SC (or UU) contrasts.

Several other non-native assimilation types have been compared in adult studies from my own and other laboratories. In a direct comparison of TC, CG and SC contrasts, I tested English listeners' discrimination with multiple natural utterances of three additional Zulu contrasts: voiced vs. voiceless lateral fricatives, voiceless aspirated vs. ejective velar stops /k/-/k'/, and plosive vs. implosive bilabial stops. A fourth non-native pair was the Tigrinya (Ethiopian) bilabial vs. alveolar ejective contrast /p'/-/t'/ (Best, 1990). The Zulu lateral fricatives were expected to assimilate to English as TC contrasts, that is, as a voiced-voiceless English fricative contrast involving the tongue tip (i.e., /z/-/s/, "zh"-"sh" or "th" in *th̲is* vs. *th̲ink*), perhaps in combination with an /l/. The Tigrinya ejectives were likewise expected to be assimilated as a TC contrast, specifically as "odd" English /p/ and /t/. The aspirated vs. ejective velar stops were expected to assimilate as a good vs. an "odd" /k/, i.e., as a CG contrast. And the plosive vs. implosive bilabials were expected to assimilate as nearly equal English /b/s. All PAM predictions were strongly supported. Nearly all subjects assimilated the contrasts as expected, according to a posttest questionnaire that asked them to describe or give English labels to recordings of each non-native category. Moreover, the levels of AXB discrimination performance were strongly associated with their assimilation patterns. That is, the Zulu and Tingrinya TC contrasts yielded excellent, near-ceiling discrimination. The Zulu CG contrast was discriminated very well, but significantly less well than the TC contrasts. The Zulu SC contrast showed the lowest discrimination, much lower than either the TC or the CG contrasts.

Two other aspects of the results from that study were consistent more generally with perceptual learning principles. First, a recency memory effect was found on the AXB discrimination trials *only* for the SC contrast (plosive-implosive bilabials). Discrimination was significantly better when X matched the B category than when it matched the A category. Second, discrimination performance on all three Zulu contrasts was significantly better for matches on the more English-like pair member. Specifically, Zulu /k/ and /b/ were perceived as more like English /k/ and /b/, respectively, than were the contrasting Zulu /k'/ and implosive bilabial, and the voiceless lateral fricative was perceived as containing an English voiceless fricative (/s/ or "sh") more consistently than the voiced cognate was perceived as containing the corresponding voiced fricative (/z/ or "zh"), even though subjects did assimilate the lateral fricatives as a TC contrast. AXB discrimination was significantly higher when the X was the more English-like /b/, /k/, or voiceless lateral fricative than when it was the less English-like implosive bilabial, /k'/ or voiced lateral fricative.

In another study, which extended the findings of MacKain, Best, and Strange (1981), we tested several PAM hypotheses by comparing categorical perception in American and Japanese listeners for three related English consonant contrasts which bear differing relations to Japanese phonology (Best & Strange, 1992). The stimuli were computer-synthesized continua for the contrasts /r/-/l/, /r/-/w/, and /w/-/y/. All three are place of articulation contrasts between approximant consonants, involving constriction gestures that are neither complete closures as in stop consonants nor critically narrow as in fricatives. The first is not a phonological contrast in Japanese, as described earlier, and was expected to show SC assimilation or UU assimilation. In the second contrast, /r/ is of course non-native for Japanese, whereas /w/ is a native category but is produced with less lip-rounding than in English. Japanese listeners should assimilate this contrast as either a CG difference within the Japanese /w/ category, or as a UC contrast with /r/ as an uncategorizable speech sound (or, less likely, as a TC contrast with a very poor Japanese /r/). The /w/-/y/ difference is a phonological contrast in Japanese as in English, although again both elements are pronounced somewhat differently in the two languages. It should therefore be assimilated as a TC contrast by Japanese listeners. Although we did not obtain posttest assimilation judgments from the Japanese listeners, the pattern of consistency in their categorization and discrimination of the three continua fits well with PAM predictions. That is, their best performance was on /w/-/y/, where they matched American listeners' performance levels; their lowest performance was on /r/-/l/, where the Americans performed as well as they did on /w/-/y/ and /w/-/r/. Those Japanese who were least experienced with English showed essentially chance performance levels on /r/-/l/ but were substantially better than chance on /w/-/r/ and especially on /w/-/y/.[9] Japanese with intensive English experience performed more similarly to Americans on /r/-/l/, as summarized earlier for MacKain et al. (1981), and also on /w/-/r/; however, there was no effect of English experience on Japanese performance with /w/-/y/.

Several adult studies from other labs are also consistent with PAM predictions, although they were not designed to test PAM. Werker and Tees (1984a) tested English speakers' discrimination of Hindi breathy voiced vs. voiceless aspirated dental stops and dental vs. retroflex voiceless unaspirated stops, as well as Nthlakampx velar-uvular ejectives /k'/-/q'/. They found listeners better able to discriminate the first contrast than the other two. This finding is consistent with PAM, given that the latter two contrasts are each likely to be assimilated as an SC contrast, specifically as /d/ and /k/, respectively. The former contrast, however, is likely to be assimilated either as /d/-/t/, a TC voicing contrast, or as a CG difference in which the Hindi breathy voiced dental is heard as a deviant English /t/. The authors had undertaken the study to test whether allophonic experience in the native language may account for variations in discriminability of different non-native contrasts (see also Werker et al., 1981; Werker & Tees, 1984b). As they note, although the allophonic explanation may be compatible with good discrimination of the Hindi dental voicing contrast (English has dental /t/ allophones) and poor discrimination of Nthlakampx ejectives (English has no ejective allophones), it is inconsistent with the poor discrimination of the Hindi dental-retroflex contrast (English does have dental allophones of /d/). Interestingly, however, a separate study found that listeners who had had experience with Hindi in their first year of life were better able than those without such experience to discriminate the dental-retroflex contrast as adults (Tees & Werker, 1984).

Two other reports have explicitly evaluated PAM hypotheses against several other possible accounts for variation in perception of differing non-native speech contrasts. One focused in depth on the Hindi dental-retroflex distinction in initial position, investigating English listeners' perception of that place of articulation contrast within each of four different voicing settings: voiced, voiceless aspirated, breathy voiced (i.e., voiced aspirated), and voiceless unaspirated (Polka, 1991). The former two voicing patterns occur for initial stops in English, whereas the latter two do not. Performance on the four place of articulation contrasts was not uniform, but rather was near chance for the former two voicing patterns, better than chance for the breathy voiced one, and better still for the voiceless unaspirated one.[10] This pattern of results led Polka to reject an account based on the lack of phonological status of the dental-retroflex stop contrast in English, as well as an account based on exposure to dental allophones of /t/-/d/ in English. An account in terms of the acoustic salience of the formant transitions in the various contrasts was also inconsistent with the observed performance pattern, given that formant transitions are most salient acoustically in the voiced dental-retroflex contrast, which was the most difficult for English listeners to discriminate. However, an assimilation account seemed to work well, in that most listeners heard both members of the poorly discriminated voiced dental-retroflex contrast as /d/ and both members of the voiceless aspirated dental-retroflex contrast as /t/, i.e., as SC contrasts. But they heard the more easily discriminated voiceless unaspirated dental-retroflex contrast as "th" (*th*is) - /d/ and breathy voiced dental-retroflex contrast as /d/-/t/, i.e., the latter two contrasts appear to have been heard as TC contrasts.

In a related study, Polka (1992) examined English and Farsi listeners' perception of the velar-uvular stop distinction in two voicing contexts: voiced (native to Farsi only) and ejective (native to neither language). On the voiced velar-uvular contrast, English listeners perceived the uvular category as "bad" exemplars of English /g/ or as no clear English consonant, thus assimilating the contrast as a CG or UC difference, which they discriminated above chance. Most listeners in both groups performed poorly on the non-native ejective contrast, describing it either in terms corresponding to an SC assimilation pattern or a UU assimilation pattern. The few subjects in both groups who showed good discrimination described the latter sounds in terms corresponding to TC, CG or UC assimilation. A separate group of English listeners showed comparable, above-chance discrimination levels on the voiced and the ejective contrast, though with a trend toward better discrimination of the Farsi voiced contrast. They described the Farsi voiced contrast in CG or UC assimilation terms and the ejective contrast in SC or UC assimilation terms. Thus, the findings from these two studies are also generally consistent with the predictions of the Perceptual Assimilation Model.

In contrast with the evidence that adults assimilate non-native contrasts with respect to native phonological categories, young infants show little or no effect of the ambient language on their perception of non-native consonants up to about 8 months of age. A number of studies have shown, however, that language-specific influences begin to appear by 8-10 months and are well-established by 10-12 months. But how closely does the 10-12 month old's discrimination of various non-native consonant contrasts mirror the pattern found in adults? In other words, are one-year olds likely to have discovered the same higher-order invariants in native speech contrasts as adults have? Have they yet discovered even that most basic aspect of the phonological component of the grammar—phonological contrast? According to the perceptual learning account of infant speech perception developed here, the answer to the last two questions should be "no."

As with the literature on adult tests of cross-language speech perception, initial reports of a decline by 10-12 months in infants' discrimination of non-native consonants used contrasts that adults from their language community assimilate as SC types. In a conditioned head-turn procedure (see Eilers, Wilson, & Moore, 1977), Werker and colleagues found that English-learning 6-8 month olds discriminate the Hindi voiceless unaspirated dental-retroflex stops, the Hindi breathy voiced vs. voiceless aspirated dental stops , and the Nthlakampx velar-uvular ejectives /k'/-/q'/. But by 10-12 months of age infants have essentially ceased to discriminate the first and third of these (the latter age was not tested on the second contrast). Hindi-learning and Nthlakampx-learning infants, of course, still discriminate their native contrasts by 10-12 months (Werker et al., 1981; Werker & Tees, 1984a). Moreover, when presented with a computer-synthesized continuum ranging from /b/ to dental to retroflex stops, 6-8 month old English-learning infants, 10-12 month old Hindi infants, and Hindi adults perceive three separate categories, whereas 10-12 month old English-learning infants and English-speaking adults hear only two categories corresponding to /b/ and d/ (Werker & Lalonde, 1988).

A recent study from my lab extended PAM directly to infants' perception of additional types of non-native assimilation types (Best et al., 1990). In this study, 6-8 month old and 10-12 month old American English-learning infants each participated in three discrimination tests with non-native consonant contrasts from Zulu, the same ones that had been used in the adult study summarized earlier (Best, 1990): plosive vs. implosive bilabial stops, voiceless aspirated vs. ejective velar stops /k/-/k'/, and voiced vs. voiceless lateral fricatives. The infants were tested using a conditioned visual fixation habituation procedure (see Best, McRoberts, & Sithole, 1988; Horowitz, 1975; Miller, 1983). As summarized earlier, English-speaking adults assimilated the lateral fricatives as a TC contrast, the velars as a CG contrast, and the bilabials as an SC contrast. Their discrimination levels followed the order TC > CG >> SC. In the infant study, the 6-8 month olds discriminated all three contrasts. The 10-12 month olds, however, failed to discriminate all three Zulu contrasts, unlike both the younger infants and the adults. The most difficult contrast for them was the lateral fricative distinction. Rather than showing even a small (non-significant) fixation increase from the end of habituation to the beginning of the test phase, as they had shown for the other two contrasts, in the lateral fricative test they simply showed a further decline, or continuation of habituation.

It is noteworthy that the TC lateral fricative contrast was especially difficult for the 10-12 month olds, given that, as a TC contrast, it was the easiest of the Zulu contrasts for adults. The older infants' difficulty might be related to the fact that most adults assimilated the lateral fricatives to various consonant clusters, many of which were not phonotactically permissible in initial position in English, such as "zhl" and "shl." In other words, the adults did not find a simple segmental contrast in English, or even a pair of permissible phonotactic sequences, to which they could assimilate the lateral fricatives. Not surprisingly, then, the older infants may have been unable to consistently detect any familiar native gestural constellations in the lateral fricatives, and may have instead perceived them as a UU assimilation type, for which discrimination is expected to be poor or perhaps as an SC assimilation type re: English (both Zulu fricatives had /l/-like properties according to many adults.[11] The older infants also failed to show significant discrimination of the velar voiceless aspirated vs. ejective /k/-/k'/, which was a fairly easy CG contrast for adults. On this contrast, they showed their largest average increase in fixation during the test phase, nearly as large as that of the 6-8 month olds, but they also showed a high degree of variability. This pattern suggests two possibilities that warrant further investigation: 1) the infants may have assimilated /k/-/k'/ as a CG contrast and shown a prototype asymmetry effect (Kuhl et al., 1992; Polka & Werker, in press) in which discrimination depended on whether they habituated to the English-like Zulu /k/ or the non-prototypical /k'/; 2) some of the infants may have assimilated /k/-/k'/as a SC contrast, failing to hear that the voicing lag in /k/ is aspirated while the lag in /k'/ completely blocks airflow (i.e., is silent), whereas others may have heard the aspiration difference and shown CG assimilation. The first possibility would result in significant test-order effects in discrimination levels, whereas the second would not.

The good discrimination of the lateral fricative and velar voicing contrasts by both 6-8 month olds and English-speaking adults, but poor discrimination by 10-12 month olds, indicates a temporary dip in development perhaps comparable to those noted earlier in the phonological properties of toddler's single word productions and in their perception of minimal contrasts in meaningful words. Thus, it may be evidence of progress in the discovery of higher-order phonological category information in speech. To examine the time-course of the transitional period for these two contrasts, Glendessa Insabella and I tested English-speaking 4 year olds, using the same conditioned fixation habituation procedure as we had with the infants (although they had to be instructed that their fixations controlled the audio, and that they should tell us afterwards whether "the sounds changed" at some point during the test) (Insabella & Best, 1990). We had to assure that this procedure was sensitive enough to detect discrimination for a contrast we knew they should be able to hear, so all children had to show fixation recovery on one test with English /b/-/d/. Because these older children would only tolerate two tests in a session, we gave one group the Zulu lateral fricative distinction as their second test; the other group got the Zulu velar voicing contrast as their second test. The 4-year-olds, unlike the 10-12 month olds, easily discriminated the /k/-/k'/ contrast. However, they still failed to discriminate the lateral fricative contrast. Thus, they had already come into line with adult performance on the CG contrast, but still showed depressed performance on the TC contrast which had proven easiest of all for the adults. The reversal of the developmental dip for the CG contrast but not for this TC contrast should not be particularly surprising, given the complexity of the adults' assimilation patterns for the latter contrast, as noted above. The prolonged difficulty with the lateral fricative contrast is to be expected according to the outline of perceptual learning discussed earlier, in that the most common assimilations for adults involved consonant clusters rather than single segments, and many of the clusters were not even permissible in initial position in English. However, adults' assimilations for /k/-/k'/ were much simpler category goodness differences for a single English segment (/k/).

It is crucial to note, in light of the preceding discussion, that 10-12 month olds do *not* fail with all non-native contrasts. In a follow-up study with 6-8 and 10-12 month olds, using the visual fixation habituation procedure but with a more stringent habituation criterion, infants completed three tests: the Zulu lateral fricatives, the Tigrinya ejective contrast /p'/-/t'/ that adults had assimilated as a TC contrast and discriminated quite well (Best, 1990), and an English fricative voicing contrast (/s/-/z/) (Best, 1991). The younger infants discriminated all three contrasts. This time the older group discriminated an adult TC contrast, the Tigrinya ejectives. But they still failed with the TC lateral fricative contrast. This failure could not be attributed to a general difficulty with fricative voicing distinctions, because they were well able to discriminate the native English /s/-/z/ contrast. Given that they could discriminate the TC ejective /p'/-/t'/ contrast that showed consistent, single-segment-based assimilation by adults, these findings lend strength to the interpretation given above for the difficulties 10-12 month olds and even 4-year olds have with the lateral fricatives.

Another study showed that older infants also clearly discriminate a non-native contrast that adults assimilate as an NA distinction, as predicted by PAM in concert with the perceptual learning approach (this was actually the first PAM study in chronological terms). Infants at 6-8, 8-10, 10-12, and also 12-14 months were tested on the Zulu click contrast on which adults had shown their "lowest" discrimination performance—still fairly high at 80% correct—the lateral vs. apical voiceless unaspirated clicks (Best et al., 1988). This study used the same conditioned fixation procedure as Best (1990). All infants also completed a test with English /b/-/d/. All four age groups clearly discriminated the click contrast, even though they could not have had even allophonic experience with such sounds in English utterances. Because we had used a rather different procedure than the head-turn procedure that Werker used in her earlier reports of a decline in 10-12 month olds' discrimination of several non-native consonant contrasts (e.g., Werker et al., 1981; Werker & Tees, 1984a), we conducted a follow-up study. Using our fixation procedure, we gave 6-8 and 10-12 month olds a test on the clicks, one on /b/-/d/, and one on the Nthlakampx velar-uvular ejective contrast /k'/-/q'/ used by Werker (Best & McRoberts, 1989). The procedural difference did not matter—Werker's findings of discrimination at 6-8 months and failure at 10-12 months for the /k'/-/q'/ contrast was replicated, as was our previous finding of continued discrimination for the Zulu clicks at both ages.

All told, then, the infant findings with non-native consonants suggest increasing sensitivity to native gestural constellations, which negatively influences 10-12 month olds' perception of many but not all non-native contrasts. However, the patterning for which non-native contrasts are discriminated by older infants, and which are not, differs in some telling ways from that of adults in their language community. Although they discriminate two contrasts that adults discriminate fairly easily to very easily—an NA contrast and a TC contrast that adults consistently assimilate to a simple segmental contrast in the native phonology—these older infants fail to discriminate two other contrasts that adults also discriminate quite easily —a CG contrast and another TC contrast that shows a more complex and somewhat idiosyncratic assimilation pattern. These findings are consistent with the possibility that one-year olds do not recognize the higher-order gestural invariants specifying *phonological relations,* including minimal phonological contrasts. The infant's detection of the somewhat lower-order invariants corresponding to native phonetic categories may not mark the emergence of true segmental phonology. Rather, the infant's detection of phonological contrast *per se* may be crucially linked to a growing awareness of word-meaning associations (see Lloyd, Werker, & Cohen, 1993), which initially reflects gestural organization at the word or phrase level rather than the segmental level (e.g., Studdert-Kennedy, 1989, 1991) As stated earlier, perception of minimal phonological contrasts in meaningful contexts may not appear until around 18-19 months (Werker & Pegg, 1992), generally coincident with the vocabulary spurt (50+ words) and primitive syntactic constructions in productive language development.

Vowel contrasts. Much less research has examined language-specific effects on adults' or infants' discrimination of vowel contrasts. However, the few available non-native vowel findings on adults are consistent with PAM predictions, excepting that thus far no vowel contrasts have met the definition of Non-Assimilable types, i.e., none are perceived as nonspeech sounds. The possibility of NA vowel contrasts, in fact, seems quite remote given the basic commonality of voicing and manner of gestures involved in vowel production. Vowels are associated with a more open vocal tract than consonants, and slower, more global gestures involving primarily the larger extrinsic muscles rather than the small intrinsic muscles of the tongue (with some concomitant jaw and lip movements) (e.g., Fowler, 1980). Vowel color is differentiated primarily by the location and height of the tongue at its closest approximation to the upper surface of the vocal tract. Vowel contrasts may also involve length (duration) and voice quality differences (e.g., creaky voice). Other differences in the production and in the phonological functions of vowels versus consonants may ultimately be important for understanding adult cross-language assimilation patterns and early developmental changes in perception of non-native contrasts (see Best, 1993). For example, vowels usually provide the sonority peaks in syllable nuclei (open airflow through vocal tract); vowels carry the prosodic properties of utterances much more than consonants do; speech errors occur among vowels or among consonants but never cross between the two classes; and articulatory movements affect the two classes in opposite manners under stress and speech rate variations (see Fowler, 1980).

Findings on English vowel perception by native Spanish-speaking adults (Flege, 1991b, in press) fit well within the PAM predictions, although the research was not motivated by the model. Spanish contains only five vowels: /i/ as in *si*, /a/ as in *casa* (more fronted than English /ɑ/), /e/ as in *mes* (roughly "ay" but not diphthongized as in English), /o/ as in *yo* (not diphthongized as in English) and /u/ as in *su* (not diphthongized as in English). It does not have "eh", "ih", /æ/ as in *bat*, "uh", short "oo" as in *book*, "aw," or several other English vowels Thus, English /ɑ/ should be assimilated by Spanish listeners as a moderately deviant exemplar of Spanish /a/. English "ih"" "eh," and /æ/ should be heard as uncategorizable vowels (with respect to each other), or perhaps as poor category exemplars with respect to Spanish /i/, /e/ and /a/, respectively. That is, English "ih"-"eh" and "eh"-/æ/ should be assimilated as UU types *vis a vis* Spanish phonology, and thus should show relatively poor discrimination, whereas /ɑ/-/æ/ may show SC or weak CG assimilation pattern and rather poor discrimination. In contrast, /i/-"eh" should show UC or TC assimilation and near-perfect discrimination, while "ih"-/i/ should likewise show UC assimilation or a strong CG difference and very good discrimination. Discrimination levels for these contrasts in a recent study by Flege (in press) are consistent with this assimilation account. All contrasts described except for /i/-"eh" were tested with native Spanish listeners. They showed very good discrimination for "ih"-/i/, and poor discrimination for the other three contrasts. The relation between discrimination performance and actual assimilation patterns cannot be determined, however, because the listeners assimilations were not assessed. Flege accounts for the findings with his Speech Learning Model, which is concerned with whether non-native sounds are "identical," "similar," or completely "new" with respect to native phonological categories (for details, see Flege, 1991b).

Also compatible with assumptions about adults' assimilation of non-native segments to their native phonology, Rochet (in press) found differences in the assimilation of the Canadian French high front-rounded vowel /y/ by Portuguese and English listeners that corresponded to differences in productions of /i/ and /u/ in those two languages. Specifically, English listeners strongly tended to assimilate French /y/ as an /u/, whereas Portuguese listeners assimilated it as an /i/. Also, Polka (submitted) found that English listeners assimilated German high front lip-rounded /y/ and high back rounded /u/ as a strong CG difference for English short "oo," and German mid-high front rounded /Y/ vs. mid-high back rounded /U/ as a weaker CG difference for short "oo." She assessed assimilation patterns directly via a keyword identification task, in which listeners had to choose from a list of words that reflected the inventory of English vowels (e.g., *hid, hoed, heed, heard,* etc.) to characterize the perceived closest match for each non-native vowel. Discrimination was very good for both German contrasts, but significantly better for /y/-/u/ than for /Y/-/U/, which Polka interpreted to be consistent with PAM's predictions.

Finally, in a recent study completed in my laboratory (Best, Faber, & Levitt, in preparation), English-speaking adults were presented with three French vowel contrasts, two Norwegian contrasts, and a Thai contrast. The non-native vowel contrasts tested were: French high front-rounded /y/ vs. mid front-rounded /œ/ were generally assimilated as the TC contrast long vs. short "oo" (*boot-book*), and French /œ/ vs. less rounded French schwa /ə/ were generally assimilated as the TC short "oo"-"uh." Both were discriminated very well. Similarly, the Norwegian high front in-rounded /ʉ/ and high front unrounded /i/ were assimilated unanimously as the TC contrast short "oo"-"ee" and were also discriminated perfectly. French /o/-/õ/ (nasalized "o") were assimilated as either a strong CG difference for English "o" or as a TC contrast (e.g., "o"-"aw") and were discriminated very well. Thai high back unrounded /ɯ/ and high mid-back unrounded /ʉ/ were assimilated as either a moderate CG difference for English "uh" or, for some subjects, to the TC contrast short "oo"-"uh" and was discriminated slightly less well than the other TC and CG contrasts. Finally, Norwegian high front out-rounded /y/ (which has less lip-rounding than French /y/: Linker, 1985) and /i/ were assimilated by nearly all subjects as comparably good /i/, that is, as a SC type; discrimination was much poorer for this contrast than for the others. When individual subjects' assimilations were grouped according to TC type vs. CG type vs. SC type, regardless of the specific non-native vowels involved, the results clearly upheld PAM predictions: discrimination was near ceiling for TC assimilations, very good but significantly lower for CG assimilations, and much lower for SC assimilations.

Three very recent findings with infants are relevant to understanding the course of perceptual learning for vowels, although only one explicitly evaluated PAM hypotheses. All three studies point to differences between vowels and consonants in the development of native-language effects on perception. In one study of 6 month olds, English-learning and Swedish-learning infants showed vowel prototype effects only for a native vowel and not for a non-native one (Kuhl et al., 1992). Comparison of this result to the vowel prototype effects found for both native and non-native vowels in English- versus Spanish-learning newborns (Walton & Socotch, 1993) suggests a developmental decline between birth and 6 months in detecting goodness-of-fit differences for unfamiliar vowel categories. This suggests that the invariants detected in native vowels by 6 month olds vs. younger infants are different, a possibility supported by a third recent finding. Both German CG vowel contrasts from the Polka, (submitted) adult study described above were discriminated by 4-1/2 month olds, who showed no asymmetry in discrimination between the more English-like and the less English-like vowel in each pair. That is, there was no vowel prototype effect on discrimination. However, by 6 months of age, infants discriminated the German vowels only if the habituation or background stimulus was a non-prototype for English (according to the adult judgments), consistent with greater generalization to the prototype than the non-prototype. By 10-12 months, discrimination of both German contrasts failed regardless of the direction of stimulus change (Polka & Werker, in press). The results provide another example of non-native contrasts that are discriminated quite well as CG contrasts by adults in the infants' language environment but which are not discriminated by infants over a certain age, the developmental pattern that was found for discrimination of Zulu /k/-k'/ (Best et al., 1990). Taken together, the infant vowel perception findings suggest that native language effects appear earlier for perceptual prototype effects for non-native vowels (around 6 months) than for discrimination of non-native consonant contrasts (around 10-12 months). The argument offered here is that infants discover relational invariants associated with native vowels earlier than higher-order invariants associated with native consonants.

Why do infants show changes in perception of non-native vowels earlier than consonants? Why does the emergence of native-language effects on vowel perception but not consonant perception precede infants' earliest word-meaning associations? Both observations suggest that the invariants infants first discover in native vowels are simpler and/or easier to detect than those discovered in native consonants. There are a number of possible reasons for this developmental asymmetry. Vowel invariants may be easier to discover because the slower vowel gestures are more stable within the flow of information and are evident over a longer period of time than consonants. Different gestural invariants may be extracted for the two classes because the style and complexity of articulatory movements differ. Vowels also carry the prosody of an utterance. Thus the information for vowel invariants may be salient to the young infant at the broader and more attention-getting prosodic level of sound structure in utterances.

Further work on language-specific attunement to speech

Generally, the findings on adults' and infants' perception of non-native segmental contrasts fit well with the Perceptual Assimilation Model and the basic principles of an ecological approach to perceptual learning of the information in native speech. However, a number of important questions remain unanswered, and must be pursued in future research. For example, we still do not know how or even whether infants actually assimilate non-native sounds to native phonetic categories. Nor do we know which features or invariants they actually extract from either native or non-native speech. Generating the methodology for assessing these issues will not be easy. Ultimately, techniques will also be needed to investigate the development of perceptual sensitivity to more abstract phonological properties such as allophonic relations, allomorphy (e.g., the voiceless vs. voiced plural marker in *cats* vs. *dogs*), and grammatical effects on phonetic forms (e.g., unreleased /t/ in *sit* vs. flap in *sitting*).

Indeed, it is still largely unknown exactly what information is captured in the invariants for adult speech perception, especially the higher-order invariants, although cross-modal speech perception research indicates that the crucial information is gestural in nature, and is not specified in purely auditory terms but rather is amodal (e.g., Fowler & Dekle, 1991; Summerfield, 1978; Walton & Bower, in press). Much more work will be needed on this issue, which should benefit from the ecological approach to speech production and its phonological organization (e.g., Browman & Goldstein, 1989, 1990a, 1992a; Kelso, Saltzman, & Tuller, 1986; Saltzman & Munhall, 1989). It seems likely that characterizing the invariants in speech perception will depend on careful mathematical and physical analyses as it has in other domains where, for example, a single parameter (termed Tau) has been mathematically determined to be the singular invariant that specifies time to contact for an observer moving toward an object (Lee, 1976; Lee, Young, & Rewt, 1992) or for a trajectile moving toward an observer (Savelsbergh, Whiting, & Bootsma, 1991; see also Michaels & Oudejans, 1992), including audible but unseen objects rolling toward a listener (Shaw, McGowan, & Turvey, 1991).

In searching out the higher-order invariants for perception of native and non-native speech, it will probably be necessary also to view the native phonology as an organized system. That is, ultimately it will be important to conceive of the perceptual effects of phonological differences between languages more comprehensively, as effects of systemic differences, and not simply differences in elements or contrasts that one language has and another lacks. This caveat is motivated by proposals that phonological systems are self-organizing, and specifically that this leads to maximal dispersion among the elements of language-specific phonological inventories (Lindblom, 1992; Lindblom, Krull, & Stark, 1993; Lindblom, MacNeilage, & Studdert-Kennedy, 1983). But even that work has not addressed how the "optimization of phonetic space" by a language might be expected to affect a listener's perception of particular non-native contrasts. However, as Lindblom points out (Lindblom, Krull, & Stark, 1993) the principle of maximal dispersion would benefit the learning of the native sound system by drastically reducing the size of the phonetic space that must be explored to discover the sound patterning of the ambient language. The relationships among elements in the system would help to illuminate precisely which differences are critical in the language, and thereby reduce the information that must be picked up subsequently by the perceiver. The Perceptual Assimilation Model is quite amenable to the conception of the phonological system as an optimization of phonetic space by a given language, but further effort is obviously needed to work out the implications in detail.

CONCLUSION

What is innate about the development of the phonological component of a language's grammar? That is, what is it that provides the constraints on acquisition of possible phonological systems? By the ecological reasoning presented in this chapter, the answer is that what is innate—what provides the constraints on phonologies and their development—is the structure and dynamic possibilities of the human vocal tract. To a first approximation, this claim is in line with the underlying assumptions of Chomsky and Halle themselves, whose universal phonetic features were initially based on articulatory concepts. The point on which I disagree with them is their assumption that the constraints are specified innately in the mind. By the ecological view proposed here, the constraints are, instead, literally in the *physical* head, in the vocal tract itself and in the lawful physical effects that its configuration and movements have on the temporally-varying shape of its acoustic product.

Chomsky and Halle (1968) were correct in suggesting that the listener who knows a language hears the phonetic shapes made familiar by experience with that language. This claim, I have argued, can be extended even to predict that the listener hears echoes of those familiar, native phonetic shapes in the non-native sounds and contrasts of unfamiliar languages. But I part ways

with their reasoning about the causal mechanisms, and about the source of listeners' knowledge. Instead, I claim that listeners hear the phonological structure of their native language in non-native speech because they have learned to detect the gestural invariants that are directly available in the information flow from the language environment. Listeners become attuned to these gestural patterns and pick up the invariants specifying those familiar patterns wherever the stimulation provides criterial evidence for them, even in non-native sounds. This attunement to native gestural invariants begins in infancy but extends over development and into adulthood, where it should even help to account for perceptual changes during the learning of additional languages.

REFERENCES

Abbs, J., & Gracco, V. (1984). Control of complex motor gestures: Orofacial muscle responses to load perturbations of the lip during speech. *Journal of Neurophysiology, 51,* 705-723.

Anderson, S. R. (1985). *Phonology in the twentieth century: Theories of rule and theories of representations.* Chicago: University of Chicago Press.

Archangeli, D. (1988). Aspects of underspecification theory. *Phonology, 5,* 183-207.

Archangeli, D., & Pulleyblank, D. (in press). *The content and structure of phonological representations.* Cambridge, MA: MIT Press.

Aslin, R. N. (1987). Visual and auditory development in infancy. In J. D. Osofsky (Ed.), *Handbook of infant development* (Vol. 1, pp. 5-97). New York: Wiley.

Aslin, R. N., Pisoni, D. B., & Jusczyk, P. W. (1983). Auditory development and speech perception in infancy. In M. Haith & J. Campos (Eds.), *Handbook of child psychology Vol. 2: Infancy and developmental psychobiology.* New York: Wiley.

Berko, J. (1958). The child's learning of English morphology. *Word, 14,* 150-177.

Bernstein Ratner, N. (1984). Patterns of vowel modification in mother-child speech. *Journal of Child Language, 11,* 557-578.

Bernstein Ratner, N. (1986). Durational cues which mark clause boundaries in mother-child speech. *Journal of Phonetics., 14,* 303-309.

Bernstein Ratner, N. (1993). Interactive influences on phonological behavior: A case study. *Journal of Child Language, 20,* 191-197.

Bernstein Ratner, N., & Luberoff, A. (1984). Cues to post-vocalic voicing in mother-child speech. *Journal of Phonetics, 12,* 285-289.

Bernstein Ratner, N., & Pye, C. (1984). Higher pitch in BT is not universal: Acoustic evidence from Quiche Mayan. *Journal of Child Language, 11,* 515-522.

Bertoncini, J., Bijeljac-Babic, R., Jusczyk, P. W., Kennedy, L. J., & Mehler, J. (1988). An investigation of young infants' perceptual representations of speech sounds. *Journal of Experimental Psychology: General, 117,* 21-33.

Best, C. T. (1984). Discovering messages in the medium: Speech and the prelinguistic infant. In H. E. Fitzgerald, B. Lester, & M. Yogman (Eds.), *Advances in pediatric psychology* (Vol. 2, pp. 97-145). New York: Plenum.

Best, C. T. (1990). Adult perception of nonnative contrasts differing in assimilation to native phonological categories. *Journal of the Acoustical Society of America, 88,* S177.

Best, C. T. (1991). Phonetic influences on the perception of nonnative speech contrasts by 6-8 and 10-12 month-olds. Presented at the meeting of the *Society for Research in Child Development*. Seattle WA, April.

Best, C. T. (1993). Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 289-304). Dordrecht, the Netherlands: Kluwer Academic Publishers.

Best, C. T. (in press a). The emergence of native-language phonological influences in infants: A perceptual assimilation model. To appear in J. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words.* Cambridge MA: MIT Press.

Best, C. T. (in press b). A direct realist view of cross-language speech perception. To appear in W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research.* Timonium, MD: York Press.

Best, C. T., Faber, A., & Levitt, A. (in preparation). Association between adults' perceptual assimilation and discrimination of diverse non-native vowel contrasts.

Best, C. T., Goodell, E., & Wilkenfeld, D. (in preparation). Phonologically-motivated substitutions in a 20-22 month old's imitations of intervocalic alveolar stops.

Best, C. T., & McRoberts, G. (1989). Phonological influences on infants' perception of two non-native speech contrasts. Presented at the *Society for Research in Child Development*, Kansas City, April.

Best, C. T. McRoberts, G. W., & Sithole, N. N. (1988). The phonological basis of perceptual loss for non-native contrasts: Maintenance of discrimination among Zulu clicks by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 345-360.

Best, C. T., McRoberts, G. W., Goodell, E., Womer, J. S., Insabella, G., Kim, P., Klatt, L., Luke, S., & Silver, J. (1990). Infant and adult perception of nonnative speech contrasts differing in relation to the listener's native phonology. Presented at meeting of the *International Conference on Infant Studies.* Montreal, April.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics, 20,* 305-330.

Bialystock, E. (1988). Levels of bilingualism and levels of linguistic awareness. *Developmental Psychology, 24,* 560-567.

Bloch, B. (1950). Studies in colloquial Japanese IV: Phonemics. *Language, 26,* 86-125.

Bohannon, MacWhinney, B., & Snow, C. (1990). No negative evidence revisited: Beyond learnability or who has what to prove to whom. *Developmental Psychology, 26,* 221-226.

Bornstein, M. H. (1979). Perceptual development: Stability and change in feature perception. In M. H. Bornstein & W. Kessen (Eds.), *Psychological development from infancy.* Hillsdale NJ: Erlbaum.

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook. 3,* 219-252.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 62,* 201-251.

Browman, C. P., & Goldstein, L. (1990a). Representation and reality: Physical systems and phonological structure. *Journal of Phonetics, 18,* 411-424.

Browman, C. P., & Goldstein, L. (1990b). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics, 18,* 299-320.

Browman, C. P., & Goldstein, L. (1990c). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 341-376). Cambridge, UK: Cambridge University Press.

Browman, C. P., & Goldstein, L. (1992a). Articulatory phonology: An overview. *Phonetica, 49,* 155-180.

Browman, C. P., & Goldstein, L. (1992b). Response to commentaries. *Phonetica, 49,* 222-234.

Carney, A. E., Widin, G. P., & Viemeister, N. F. (1977). Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America, 62,* 961-970.

Catford, J. C. (1988). *A practical introduction to phonetics.* Oxford: Clarendon Press.

Chambers, J. K. (1992). Dialect acquisition. *Language, 68,* 673-705.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge MA: MIT Press.

Chomsky, N. (1972). *Language and mind.* Cambridge MA: Harcourt Brace Jovanovich, Inc.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English.* New York: Harper & Row.

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (submitted). Do infants perceive word boundaries? An empirical approach to the bootstrapping problem for lexical acquisition.

Clements, G. N. (1985). The geometry of phonological features. *Phonology Yearbook, 2,* 225-252.

Clements, G. N. (1992). Phonological primes: Features or gestures? *Phonetica, 49,* 181-193.

Cohn, A. (1990). Phonetic and phonological rules of nasalization. *UCLA Working Papers, 76,* May.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development, 61,* 1584-1595.

Costa, P., & Mattingly, I. G. (1981). Production and perception of phonetic contrast during phonetic change. *Journal of the Acoustical Society of America, 69,* S67 (abstract).

Cruttenden, A. (1974). An experiment involving comprehension of intonation in children from 7 to 10. *Journal of Child Language, 1,* 221-231.

Dent, C. H. (1990). An ecological approach to language development: An alternative functionalism. *Developmental Psychobiology, 23,* 679-703.

Dent, C., & Rader, N. (1979). Perception, meaning and research in semantic development. In P. French (Ed.), *The development of meaning: Pedolinguistic series* (pp. 178-230). Japan: Bunka Hyoron.

Derwing, B. (1973). *Transformational grammar as a theory of language acquisition.* Cambridge, UK: Cambridge University Press.

Di Paolo, M. (1992). Hypercorrection in response to apparent merger of /ɑ/ and /ɔ/ in Utah English. *Language & Communication, 12,* 267-292.

Di Paolo, M., & Faber, A. (1991). Phonation differences and the phonetic content of the tense-lax contrast in Utah English. *Language Variation and Change, 2,* 155-204.

Eilers, R. E. (1977). Context-sensitive perception of naturally produced stop and fricative consonants by infants. *Journal of the Acoustical Society of America, 61,* 1321-1336.

Eilers, R. E., Gavin W. J., & Oller, D. K. (1982). Cross linguistic perception in infancy: Early effects of linguistic experience. *Journal of Child Language, 9,* 289-302.

Eilers, R. E., & Minifie, F. D. (1975). Fricative discrimination in early infancy. *Journal of Speech and Hearing Research, 18,* 158-167.

Eilers, R. E., Wilson, W. R., & Moore, J. M. (1977). Developmental changes in speech discrimination in infants. *Journal of Speech and Hearing Research, 20,* 766-780.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science, 171,* 303-306.

Elman, J. L., Diehl, R. L., & Buchwald, S. E. (1977). Perceptual switching in bilinguals. *Journal of the Acoustical Society of America, 62,* 971-974.

Faber, A. (1992). Articulatory variability, categorical perception, and the inevitability of sound change. In G. W. Davis & G. K. Iverson (Eds.), *Explanation in historical linguistics* (pp. 59-75). Amsterdam: John Benjamins Publishing Co.

Faber, A., Best, C. T., & Di Paolo, M. (1993). Cross-dialect perception of nearly-merged forms. Presented at the meeting of the Linguistic Society of America. Los Angeles, CA, January.

Faber, A., Di Paolo, M., & Best, C. T. (submitted). Perceiving the unperceivable: The acquisition of nearly-merged forms.

Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language, 15,* 419-439.

Ferguson, C. A. (1986). Discovering sound units and constructing sound systems: It's child's play. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability of speech processes* (pp. 36-53). Hillsdale, NJ: Erlbaum.

Fernald, A. (1984). The perceptual and affective salience of mother's speech to infants. In L. Feagans, C. Garvey, & R. Golinkoff (Eds.), *The origins and growth of communication* (pp. 5-29). Norwood, NJ: Ablex.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development, 8,* 181-195.

Fernald, A., & Kuhl, P. K. (1982). [Discrimination of five-syllable sequences (galas*a*laga versus gala*ta*laga) by 6 month old infants]. Unpublished findings, cited in Karzon, R. G. (1985)

Fernald, A., & Kuhl, P. K. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development, 10,* 279-293.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology, 27,* 209-221.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology, 20,* 104-113.

Fernald, A., Taeschner, T., Dunn, J, Papoušek, M., Boysson-Bardies, B., & Fukui, I. (1990). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language, 16,* 477-502.

Flege, J. E. (1984). The detection of French accent by American listeners. *Journal of the Acoustical Society of America, 76,* 692-707.

Flege, J. E. (1987). A critical period for learning to pronounce foreign languages? *Applied Psycholinguistics, 8,* 162-177.

Flege, J. E. (1989). Chinese subjects' perception of the word-final English /t/-/d/ contrast: Before and after training. *Journal of the Acoustical Society of America, 86,* 1684-1697.

Flege, J. E. (1991a). Perception and production: The relevance of phonetic input to L2 phonological learning. In T. Huebner & C. Ferguson (Eds.), *Cross-currents in second language acquisition and linguistic theory.* Philadelphia: John Benjamins.

Flege, J. E. (1991b). Orthographic evidence for the perceptual identification of vowels in Spanish and English *Quarterly Journal of Experimental Psychology, 43,* 701-731.

Flege, J. E. (in press). Second-language speech learning: Theory, findings, and problems. To appear in W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues. Timonium,* MD: York Press.

Flege, J. E., & Eefting, W. (1987). The production and perception of English stops by Spanish speakers of English. *Journal of Phonetics, 15,* 67-83.

Flege, J. E., & Fletcher, J. (1992). Listener and talker effects on the perception of degree of foreign accent. *Journal of the Acoustical Society of America, 91,* 370-389.

Fodor, J. A., Garrett, M. F., & Brill, S. L. (1975). Pi Ka Pu: The perception of speech sounds by pre-linguistic infants. *Perception & Psychophysics, 18,* 74-78.

Fourakis, M., & Port, R. (1986). Stop epenthesis in English. *Journal of Phonetics, 14,* 197-221.

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics, 8,* 113-133.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics, 14,* 3-28.

Fowler, C. A. (1989). Real objects of speech perception: A commentary on Diehl and Kluender. *Ecological Psychology, 1,* 145-160.

Fowler C. A. (1991). Sound-producing sources as objects of perception: Rate normalization and nonspeech perception. *Journal of the Acoustical Society of America, 88,* 1236-1249.

Fowler, C. A., Best, C. T., & McRoberts, G. W. (1990). Young infants' perception of liquid coarticulatory influences on following stop consonants. *Perception & Psychophysics, 48,* 559-570.

Fowler, C. A., & Dekle, D. J. (1991). Listening with eye and hand: Cross-modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 816-828.

Fowler, C. A., Rubin, P. Remez, R. E., & Turvey, M. T. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Speech production* (pp. 373-420). New York: Academic Press.

Fowler, C. A., & Smith, M. (1986). Speech perception as "vector analysis": An approach to the problems of segmentation and invariance. In J. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 123-139). Hillsdale, NJ: Erlbaum.

Friederici, A. D., & Wessels, J. M. I. (in press). Phonotactic knowledge and its use in infant speech perception. *Perception & Psychophysics.*

Gerken, L. A., Landau, B., & Remez, R. E. (1990). Function morphemes in young children's speech perception and production. *Developmental Psychology, 27,* 204-216.

Gerken, L., & McIntosh, B. J. (1993). Interplay of function morphemes and prosody in early language. *Developmental Psychology, 29,* 448-457.

Gibson, E. J. (1963). Perceptual learning. *Annual Review of Psychology, 14,* 29-56.

Gibson, E. J. (1966) Perceptual development and the reduction of uncertainty. In *Proceedings of the 18th International Congress of Psychology,* 7-17.

Gibson, E. J. (1969). *Principles of perceptual learning and development.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

Gibson, E. J. (1971). Perceptual learning and the theory of word perception. *Cognitive Psychology, 2,* 351-368.

Gibson, E. J. (1977). How perception really develops: A view from outside the system. In D. LaBerge & S. J. Samuels (Eds.), *Basic processes in Reading: Perception and Comprehension.* Hillsdale, NJ: Erlbaum Associates.

Gibson, E. J. (1988). Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge. *Annual Review of Psychology, 39,* 1-49.

Gibson, E. J. (1991). *An odyssey in learning and perception.* Cambridge, MA: Bradford Books (MIT Press).

Gibson, E. J., Gibson, J. J. (1972). The senses as information-seeking systems. *(London) Times Literary Supplement, June 23,* 711-712.

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Boston, MA: Houghton-Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston, MA: Houghton-Mifflin.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review, 62,* 32-41.

Gillette, S. (1980). Contextual variation in the perception of L and R by Japanese and Korean speakers. *Minnesota Papers in Linguistics and the Philosophy of Language, 6,* 59-72.

Gleitman, L., Gleitman, H., Landau, B., & Wanner, E. (1988). Where learning begins: Initial representations for language learning. In F. Newmeyer (Ed.), *The Cambridge Linguistic Survey.* Cambridge MA: MIT Press.

Goldring (Zukow), P. (1991). Early steps toward language: How social affordances educate attention. Presented at meeting of the *International Conference on Event Perception and Action.* Amsterdam, the Netherlands, August.

Goldsmith, J. (1976). *Autosegmental phonology.* Unpublished doctoral dissertation, MIT University.

Goldstein, L., & Browman, C. P. (1986). Representation of voicing contrasts using articulatory gestures. *Journal of Phonetics, 14,* 339-342.

Goodell, E., & Studdert-Kennedy, M. (1990). From phonemes to words or words to phonemes: How do children learn to talk? Presented at meeting of the *International Conference on Infant Studies,* Montreal, April.

Goodsitt, J. V., Morse, P. A., Ver Hoeve, J. N., & Cowan, N. (1984). Infant speech recognition in multisyllabic utterances. *Child Development, 55,* 903-910.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R." *Neuropsychologia, 9,* 317-323.

Grieser, D. L., & Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental Psychology, 24, (1),* 14-20.

Grieser, D. L., & Kuhl, P. K. (1989). Categorization of speech in infants: Support for speech-sound prototypes. *Developmental Psychology, 25,* 577-588.

Hillenbrand, J. (1983). Perceptual organization of speech sounds by infants. *Journal of Speech and Hearing Research, 26,* 268-282.

Hillenbrand, J. (1984). Speech perception by infants: Categorization based on nasal consonant place of articulation. *Journal of the Acoustical Society of America, 75,* 1613-1622.

Hirsh-Pasek, K., Kemler Nelson, D., Jusczyk, P., Wright Cassidy, K., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition, 26,* 269-286.

Hockett, C. F. (1963). The problem of universals in language. In J., H., Greenebrg (Ed.), *Universals in language.* Cambridge MA: MIT Press.

von Hofsten, C. (1980). Predictive reaching for moving objects by human infants. *Journal of Experimental Child Psychology, 30,* 369-382.

Hohne, E. A., & Jusczyk, P. W. (1992). Allophonic variation and word segmentation in infant speech perception. Paper presented at the *International Conference on Infcnt Studies,* Miami FL, May.

Holm, J. A. (1988). *Pidgins and creoles. Volume I: Theory and structure.* Cambridge, UK: Cambridge University Press.

Holmberg, T. L., Morgan, K. A., & Kuhl, P. K. (1977). Speech perception in early infancy: Discrimination of fricative consonants. Paper presented at the meeting of *The Acoustical Society of America,* Miami Beach FL (December).

Horowitz, F. D. (1975). Visual attention, auditory stimulation, and language discrimination in infants. *Monographs of the Society for Research in Child Development, 39,* Serial # 159.

Inkelas, S., & Leben, W. (1990). Where phonology and phonetics intersect. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 341-376). Cambridge, UK: Cambridge University Press.

Insabella, G., & Best, C. T. (1990). Four-year-olds' perception of nonnative contrasts differing in phonological assimilation.

Presented at meeting of the *Acoustical Society of America*, San Diego, November.

Ito, J. (1986). Syllable theory in prosodic phonology. Unpublished doctoral dissertation, University of Massachusetts, Amherst.

Jakobson, R., Fant, G., & Halle, M. (1963). *Preliminaries to speech analysis*. Cambridge MA: MIT Press.

Jakobson, R., & Halle, M. (1957). Phonology in relation to phonetics. In L. Kaiser (Ed.), *Manual of phonetics* (pp. 215-251). Amsterdam: North Holland.

Jamieson & Morosan (1986). Training non-native speech contrasts in adults: Acquisition of the English /ð/-/θ/ contrast by francophones. *Perception & Psychophysics, 40,* 205-215.

Jeffreys, W. H., & Berger, J. O. (1992). Ockham's razor and Bayesian analysis. *American, Scientist, 80,* 64-72.

Jusczyk, P. W. (1992). Developing phonological categories from the speech signal. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 17-64). Timonium, MD: York Press.

Jusczyk, P. W. (1993). Sometimes it pays to look back before you leap ahead. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 289-304). Dordrecht, the Netherlands: Kluwer Academic Publishers.

Jusczyk, P. W. (1994). Infant speech perception and the development of the mental lexicon. In J. Goodman & H. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 227-270). Cambridge: MIT Press,

Jusczyk, P. W., Bertoncini, J., Bijeljac-Babic, R., Kennedy, L., & Mehler, J. (1990). The role of attention in speech perception by infants. *Cognitive Development,*

Jusczyk, P. W., Charles-Luce, J., & Luce, P. A. (submitted). Infants' sensitivity to high frequency vs. low frequency phonetic sequences in the native language.

Jusczyk, P W., Copan, H., & Thompson, E. (1978). Perception by 2-month-old infants of glide contrasts in multisyllabic utterances. *Perception & Psychophysics, 24,* 515-520.

Jusczyk, P. W., Cutler, A., & Redanz, N. (1993). Preference for the predominant stress patterns of English words. *Child Development, 64,* 675-687.

Jusczyk, P. W., & Derrah, C. (1987). Representation of speech sounds by young infants. *Developmental Psychology, 23,* 648-654.

Jusczyk, P. W., Friederici, A. D., Wessels, J., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound structure of native language words. *Journal of Memor y and Language, 32,* 402-420.

Jusczyk, P. W., & Kemler Nelson, D. G. (in press). Syntactic units, prosody, and psychological reality during infancy. To appear in J. L. Morgan & K. D. Demuth (Eds.), *Signal to syntax.* Hillsdale NJ: Erlbaum.

Jusczyk, P. W., Pisoni, D. B., & Mullinex, J. (1992). Some consequences of stimulus variability on speech processing by two-month-olds. *Cognition, 43,* 253-291.

Jusczyk, P. W., & Thompson, E. (1978). Perception of a phonetic contrast in multisyllabic utterances by 2-month-old infants. *Perception & Psychophysics, 23,* 105-109.

Kahn, D. (1980). *Syllable-based generalizations in English phonology.* New York: Garland Press.

Karzon, R. G. (1985). Discrimination of polysyllabic sequences by one-to four-month-old infants. *Journal of Experimental Child Psychology, 39,* 326-342.

Keating, P. A. (1988). The phonology-phonetics interface. In F. Newmeyer (Ed.), *Linguistics: The Cambridge survey. Vol. 1: Grammatical theory* (pp. 281-302). Cambridge, UK: Cambridge University press.

Keating, P. A. (1990). Phonetic representations in a generative grammar. *Journal of Phonetics, 18,* 321-334.

Kelso, J. A. S., Saltzman, E. L., & Tuller, B. (1986). The dynamical perspective on speech production: Data and theory. *Journal of Phonetics, 14,* 29-59.

Kemler Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W., & Wright-Cassidy, K. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language, 16,* 55-68.

Kenstowicz, M., & Kisseberth, C. (1979). *Generative phonology.* New York: Academic Press.

Kent, R. D., Carney, P. J., & Severeid, L. R. (1974). Velar movement and timing: Evaluation of a model for binary control. *Journal of Speech and Hearing Research, 17,* 175-177.

Krakow, R., Beddor, P. S., Goldstein, L., & Fowler, C. A. (1988). Coarticulatory influences in the perceived height of nasal vowels. *Journal of the Acoustical Society of America, 83,* 1146-1158.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1982). On the control and coordination of naturally developing systems. In J. A. S. Kelso & J. E. Clark (Eds.), *The development of movement control and coordination* (pp. 5-78). Chichester, UK: John Wiley.

Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowels. *Journal of the Acoustical Society of America, 66,*1668-179.

Kuhl, P. K. (1980). Perceptual constancy for speech-sound categories in early infancy. In G. H. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child phonology: Vol. 2: Perception* (pp. 41-66). New York: Academic Press.

Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development, 6,* 263-285.

Kuhl, P. K. (1987). Perception of speech and sound in early infancy. In P. Salapatek & L. Cohen (Eds.), *Handbook of infant perception* (Vol. 2, pp. 275-382). New York: Academic Press.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics, 50,* 93-107.

Kuhl, P. K., & Miller, J. D. (1982). Discrimination of auditory target dimensions in the presence or absence of variation in a second dimension by infants. *Perception & Psychophysics, 31,* 279-292.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255,* 606-608.

Labov, W. (1974). On the use of the present to explain the past. In *Proceedings of the 11th International Congress of Linguists* (pp. 825-852). Bologna: Società Editrice Il Mulino.

Labov, W., Karen, M., & Miller, C. (1991). Near mergers and the suspension of linguistic contrast. *Language Variation and Change, 3,* 33-74.

Labov, W., Yaeger, M., & Steiner, R. C. (1972). A quantitative study of sound change in progress. In *Report on NSF contract GS-33287.* Philadelphia: U. S. Regional Survey.

Ladefoged, P. (1982). *A course in phonetics.* New York: Harcourt-Brace-Jovanovich.

Leben, W. (1978). The representation of tone. In V. Fromkin (Ed.), *Tone: A linguistic survey.* New York: Academic Press.

Lee, D. N. (1976). A theory of visual control of braking based on information about time-to-collision. *Perception, 5,* 437-459.

Lee, D. N., Young, D. S., & Rewt, D. (1992). How do somersaulters land on their feet? *Journal of Experimental Psychology: Human Perception and Performance, 18,* 1195-1202.

Levitt, A., Jusczyk, P., Murray, J., & Carden, G. (1988). Context effects in two-month-old infants' perception of labiodental/interdental fricative contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 361-368.

Liberman, A. M. (1992). The relation of speech to reading and writing. in R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning.* Amsterdam: Elsevier Science Publishers B. V.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*, 1-36.

Lindau, M. (1984). Phonetic differences in glottalic consonants. *Journal of Phonetics, 12*, 147-155.

Lindblom, B. (1992). Phonological units as adaptive emergents of lexical development. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 131-164). Timonium, MD: York Press.

Lindblom, B., Krull, D., & Stark, J. (1993). Phonetic systems and phonological development. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, and J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 399-409). Dordrecht, the Netherlands: Kluwer Academic Publishers.

Lindblom, B., MacNeilage, P., & Studdert-Kennedy, M. (1983). Self-organizing processes and the explanation of phonological universals. In B. Butterworth, B. Comrie, & O. Dahl (Eds.), *Universals Workshop* (pp. 181-203). The Hague: Mouton.

Linker, W. (1985). A cross-linguistic study of lip position in vowels. *UCLA Working Papers in Phonetics, 51*, 1-35.

Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments on comparative phonetics. *Proceedings of the 6th International Congress of Phonetic Sciences*. Prague: Academia.

Lloyd, V. L., Werker, J. F., & Cohen, L. B. (1993). Age changes in infants' ability to associate words with objects. Presented at meeting of *Society for Research in Child Development*. New Orleans, LA, March.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89*, 874-886.

MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics, 2*, 369-390.

Macken, M. A. (1992). Where's phonology? In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 249-269). Timonium, MD: York Press.

Macken, M., & Ferguson, C. S. (1983). Cognitive aspects of phonological development: Model, evidence, and issues. In K. E. Nelson (Ed.), *Children's language* (pp. 255-282). Hillsdale, NJ: Lawrence Erlbaum Associates.

Madore, B. F., & Freeman, W. L. (1987). Self-organizing structures. *American Scientist, 75*, 252-259.

Malsheen, B. J. (1980). Two hypotheses for phonetic clarification in the speech of mothers to children. In G. H. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child Phonology* (Vol. 2, pp. 173-184). New York: Academic Press.

Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics, 28*, 407-412.

Mann, V. A. (1986). Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of "l" and "r." *Cognition, 24*, 169-196.

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., & Fei, X. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development, 57 (4)*. serial # 228.

Martin, C. S., Mullennix, J. W., Pisoni, D. B., & Summers, W. V. (1989). Effects of talker variability on recall of spoken lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 676-684.

McCarthy, J. (1986). OCP effects: Gemination and anti-gemination. *Linguistic Inquiry, 17*, 207-263.

McCarthy, J. (1988). Feature geometry and dependency. A review. *Phonetica, 43*, 84-108.

McCarthy, J. (1989). Linear order in phonological representation. *Linguistic Inquiry, 20*, 71-99.

McCune, L. (1992). First words: A dynamic systems view. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 313-336). Timonium, MD: York Press.

McCune, L., & Vihman, M. (1987). Vocal motor schemes. *Papers and Reports in Child Language Development, 26*, 72-79.

McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748.

McLaughlin, B. (1978). Second-language acquisition in childhood. Hillsdale NJ: Erlbaum.

Menn, L. (1986). Language acquisition, aphasia and phonotactic universals. In F. R. Eckman, E. A. Moravcsik, & J. R. Wirth (Eds.), Markedness. (pp. 241-255). New York: Plenum Press.

Menn, L., & Matthei, E. (1992). The "two-lexicon" account of child phonology: Looking back, looking ahead. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 211-247). Timonium, MD: York Press.

Michaels, C. F., & Oudejans, R. R. D. (1992). The optics and actions of catching fly balls: Zeroing out optical acceleration. *Ecological Psychology, 4*, 199-222.

Miller, C. L. (1983). Developmental changes in male-female voice classification by infants. *Infant Behavior and Development, 6*, 313-330.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics, 18*, 331-340.

Mochizuki, M. (1981). The identification of /r/ and /l/ in natural and synthesized speech. *Journal of Phonetics, 9*, 283-303.

Mohanan, K. P. (1986). *The theory of lexical phonology*. Boston: D. Reidel Publishing Company.

Mohanan, K. P. (1992). Emergence of complexity in phonological development. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 635-662). Timonium, MD: York Press.

Morgan, J. L. (1990). Input, innateness, and induction in language acquisition. *Developmental Psychobiology, 23*, 661-678.

Newmeyer, F. J. (1980). *Linguistic theory in America: The first quarter-century of transformational generative grammar*. New York: Academic Press.

Nittrouer, S., Studdert-Kennedy, M., & McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research, 32(1)*, 120-132.

Ohala, J. J. (1990). There is no interface between phonology and phonetics: A personal view. *Journal of Phonetics, 18*, 153-171.

Oller, D. K. (1980). The emergence of the sounds of speech in infancy. In G. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child phonology, Volume 1: Production*. New York: Academic Press.

Oller, D. K., & Lynch, M. P. (1992). Infant vocalization and innovations in infraphonology: Toward a broader theory of development and disorders. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 509-536). Timonium, MD: York Press.

Oshika, B. T., Zue, V. W., Weeks, R. V., Neu, H., & Aurbach, J. (1975). The role of phonological rules in speech understanding research. *IEEE Transaction on Acoustics, Speech and Signal processing, 23*, 104-112.

Pierrehumbert, J. B. (1990). Phonological and phonetic representation. *Journal of Phonetics, 18*, 375-394.

Pierrehumbert, J. B., & Beckman, M. E. (1988). *Japanese tone structure*. Cambridge, MA: MIT Press (Linguistic Inquiry Monograph Series, 15).

Pierrehumbert, J. B., & Pierrehumbert, R. T. (1990). On attributing grammars to dynamical systems. *Journal of Phonetics, 18*, 465-477.

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessey, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 297-314.

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America, 89*, 2961-2977.

Polka, L. (1992). Characterizing the influence of native experience on adult speech perception. *Perception & Psychophysics, 52*, 37-52.

Polka, L. (submitted). Linguistic influences in adult perception of non-native vowel contrasts.

Polka, L., & Werker, J. F. (in press). Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception & Performance.*

Price, P. J. (1981). A cross-linguistic study of flaps in Japanese and in American English. Unpublished doctoral dissertation. University of Pennsylvania.

Prigogine, I. (1980). *From being to becoming: Time and complexity in the physical sciences.* San Francisco: W. H. Freeman & Co.

Prigogine, I., & Stengers, I. (1984). Order out of chaos: Man's new dialogue with nature. Toronto: Bantam Books.

Prince, A., & Smolensky, P. (1993). Optimality theory, Technical Reports of the Rutgers University Center for Cognitive Science, TR-2.

Pruitt, J. S., Strange, W., Polka, L., & Aguilar, M. C. (1990). Effects of category knowledge and syllable truncation during auditory training on Americans' discrimination of Hindi retroflex-dental contrasts. Presented at meeting of the Acoustical Society of America. State College, PA, May.

Quine, W. V. (1960). *Word and object.* Cambridge MA: MIT Press.

van Reenen, P. (1982). *Phonetic feature definitions: Their integration into phonology and their relation to speech. A case study of the feature nasal.* Dordrecht: Foris Publications.

Rochet, B. (in press). [Cross-language studies of vowel perception and production]. To appear in W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues.* Timonium, MD: York Press.

Romaine, S. (1988). *Pidgin and creole languages.* London: Longman.

Rosenblum, T., & Pinker, S. A. (1983). Word magic revisited: Monolingual and bilingual children's understanding of the word-object relationship. Child Development, 54, 773-780.

Sagey, E. C. (1986). On the ill-formedness of crossing association lines. Linguistic Inquiry, 19, 109-118.

Saltzman, E. L., & Kelso, J. A. S. (1987). Skilled actions: A task dynamic approach. Psychological Review, 94, 84-106.

Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. Ecological Psychology, 1, 333-382.

Sampson, G. (1980) *Schools of linguistics.* Stanford CA: Stanford University Press.

de Saussure, F. (1959). *Course in general linguistics.* New York: McGraw-Hill (translation of *Cours de linguistique générale.* 1916. Paris: Payot).

Savelsbergh, G. J. P., Whiting, H. T. A., & Bootsma, R. J. (1991). Grasping tau. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 315-322.

Schöner, G., & Kelso, J. A. S. (1988). Dynamic pattern generation in behavioral and neural systems. *Science, 239*, 1513-1520.

Shaw, B. K., McGowan, R. S., & Turvey, M. T. (1991). An acoustic variable specifying time-to-contact. *Ecological Psychology, 3*, 253-261.

Shaw, R., McIntyre, M., & Mace, W. (1974). The role of symmetry in event perception. In R. B MacLeod & H. L. Pick, Jr. (Eds.), *Perception: Essays in honor of James J. Gibson* (pp. 276-310). Ithaca, NY: Cornell University Press.

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics, 3*, 243-261.

Silverman, D. (1992). Multiple scansions in loanword phonology: Evidence from Cantonese. *Phonology, 9*, 289-328.

Stark, R. E. (1980). Stages of speech development in the first year of life. In G. H. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child phonology: Vol. 1: Production.* (pp. 73-92). New York: Academic Press.

Steriade, D. (1990). Gestures and autosegments: Comments on Browman and Goldstein's "Gestures in articulatory phonology." In J. Kingston & M. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech.* Cambridge UK: Cambridge University Press.

Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David, Jr. & P. B. Denes (Eds.), *Human communication: A unified view* (pp. 51-66). New York: McGraw-Hill.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics, 17*, 3-45.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training in the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics, 36*, 131-145.

Studdert-Kennedy, M. (1986). Development of the speech perceptuomotor system. In B. Lindblom & R. Zetterstrom (Eds.), *Precursors of early speech.*(pp. 205-218). New York: Stockton Press.

Studdert-Kennedy, M. (1987). The phoneme as a perceptuomotor structure. In A. Allport, D. MacKay, W. Prinz, & E Scheerer (Eds.), *Language, perception and production* (pp. 67-84). New York: Academic Press.

Studdert-Kennedy, M. (1989). The early development of phonological form. In C. von Euler, H. Forssberg & H. Lagercrantz (Eds.), *Neurobiology of early infant behavior* (pp. 287-301).Basingstoke, England: MacMillan.

Studdert-Kennedy, M. (1991). Language development from an evolutionary perspective. In N. Krasnegor, D. Rumbaugh, R. Schiefelbusch & M. Studdert-Kennedy (Eds.), *Language acquisition: Biological and behavioral determinants* (pp. 5-28). Hillsdale, NJ: Erlbaum Associates.

Studdert-Kennedy, M. (1993). Some theoretical implications of cross-modal research in speech perception. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 461-466). Dordrecht, the Netherlands: Kluwer Academic Publishers.

Summerfield, A. O. (1978). Perceptual learning and phonetic perception. In *Interrelations of the communicative senses. Proceedings of the NSF Conference at Asilomar. Washington,* DC: NSF Publications.

Tees, R. C., & Werker, J. F. (1984). Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology, 38*, 579-590.

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by adults and infants. *Child Development, 47*, 466-472.

Treiman, R., Cassar, M., & Zukowski, A. (submitted). What types of linguistic information do children use in spelling?: The case of flaps.

Turvey, M. T. (1980). Clues from the organization of motor systems. In U. Bellugi & M. Studdert-Kennedy (Eds.), *Signed and spoken language: Biological constraints on linguistic form* (pp. 41-56). Weinheim, FRG: Verlag Chemie.

Turvey, M. T. (1990). Coordination. *American Psychologist, 45,* 938-953.

Vance, T. (1987). *An introduction to Japanese phonology.* Albany, NY: State University of New York Press.

Vihman, (1992). Early syllables and the construction of phonology. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.),*Phonological development: Models, research, implications* (pp. 393-422). Timonium, MD: York Press.

Walton, G., & Bower, T. G. R. (1993). Amodal representation of speech in infants. *Infant Behavior and Development, 16,* 233-243.

Walton, G., & Socotch, T. (1993). Human newborns show a "perceptual magnet effect" for native and non-native language prototypes. Presented at meeting of the *Society for Research in Child Development.* New Orleans, March.

Werker, J. F. (1989). Becoming a native listener. *American Scientist, 77,* 54-59.

Werker, (1991). The ontogeny of speech perception. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception* (pp. 91-110). Hillsdale, NJ: Erlbaum Associates.

Werker, J. F., & Baldwin, D. A. (1991). Speech perception and lexical acquisition. Presented at meeting of the *Society for Research in Child Development,* Seattle WA, April.

Werker, J. F., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development, 52,* 349-355.

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology, 24,* 672-683.

Werker, J., & Logan, J. (1985). Cross-language evidence for three factors in speech perception. *Perception and Psychophysics, 37,* 35-44.

Werker, J. F., & McLeod, (1989). Infant preference for both male and female infant directed talk: A developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology, 43,* 230-246.

Werker, J. F., & Pegg, (1992). In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 131-164). Timonium, MD: York Press.

Werker, J. F., & Tees, R. C. (1984a). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America, 75,* 1866-1878.

Werker, J. F., & Tees, R. C. (1984b). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7,* 49-63.

Whalen, D. H. (1983). Vowel information in postvocalic fricative noises. *Language & Speech, 26,* 91-100.

Williams, L. (1979). The modification of speech perception and production in second language learning. *Perception & Psychophysics, 26,* 95-104.

Yamada, R. A., & Tohkura, Y. (1991). Age effects on acquisition of non-native phonemes: Perception of English /r/ and /l/ for native speakers of Japanese. In *Proceedings of the 12th International Congress of Phonetic Sciences, Vol. 4* (pp. 450-453). Aix-en-Provence, France: University of Aix Press.

Zukow, P., & Schmidt, C. (1988). Socializing attention: Perceptual bases for language socialization. Presented at meeting of the *International Conference on Infant Studies.* Washington DC, April.

FOOTNOTES

[*] To appear in C. Rovee-Collier & L. Lipsitt (Eds.), *Advances in infancy research.* Ablex Publishers (1994).

[†] Also Wesleyan University.

[1] Exceptions are extremely rare. For example, Native Hawaiian lacks /t/, including instead only /p/ and /k/ for its non-nasal stop consonants.

[2] Although loan word pronunciations can be affected by spelling in both donor and recipient languages, the association between spelling and pronunciation is generally not arbitrary but reflects phonological principles. The degree of transparency between spelling and pronunciation differs among languages, however, e.g., Spanish spelling is quite transparent while English spelling is much less so.

[3] The written form is another type of direct evidence that speaker-listeners can present to one another, but it is subject to at least the same limitations as the spoken form. Presumably, the evidence it carries about the underlying grammar would also be considered inadequate. In any event, normal children learn to read and write only after they have learned to talk, so the written form would generally not offer an alternative basis for language learning (see also Liberman, 1992).

[4] In fact, the relation between the individual speaker-hearer's grammatical knowledge (linguistic competence), the same speaker-hearer's actual language behavior, (linguistic performance), and the community's shared language is a complex issue. Although the matter cannot be explicated here, the reader wishing further information is referred to, e.g., Chomsky (1968; 1972), Newmeyer (1980), Sampson (1980), and de Saussure (1959).

[5] Indeed, how could one define "similar enough" if the utterances that serve as the only direct interface between different individuals' grammars inadequately reflect those grammars, and thus are by definition inadequate to validate or reliably compare them?

[6] Currently, the model assumes that articulator movement is modelled fairly well by the dynamic regime of a "point attractor," or damped mass spring, model with constant mass for each articulator. Such dynamic regimes characterize the pattern of movement of a physical system moving smoothly toward a single target ("attractor").

[7] For multilingual listeners, there may also be diachronic variations associated with code-switching, i.e., shifting from use of one language to another may effect changes in which gestural invariants are detected in an unfamiliar phonetic pattern (e.g., Elman et al., 1977; Williams, 1977).

[8] This claim should also apply to the phonological inventories of other languages, for fluent multilinguals who learned their languages during childhood. That is, childhood-onset multilinguals may be able to assimilate unfamiliar non-native sounds to categories in any of their multiple languages. Indeed, they may have greater overall sensitivity to the phonetic properties of unfamiliar phonological categories, to the extent that early learning of more than one language grants increased recognition of the arbitrariness of linguistic categories, although this sort of metalinguistic advantage has thus far been argued only for semantic and syntactic knowledge, support has been mixed (e.g., Bialystock, 1988; Rosenblum & Pinker, 1983; see McLaughlin, 1978).

[9] In addition, we found that both language groups heard a third, intermediate category between *rock* and *wok.* Tests with a second group of American listeners confirmed our suspicion that this category was clearly heard as an /l/, which falls between /w/ and "y" in place of articulation. See Best and Strange (1992) for further discussion.

[10] It should be noted that Polka used a more sensitive discrimination task, i.e. one with lower memory demands, than had Werker & Tees (1984a), which may well account for the discrepancy between the two studies in listeners' difficulty with this particular contrast.

[11] This is a new interpretation, which better handles the full array of findings than the preliminary interpretation offered in Best (in press a).

Haskins Laboratories Status Report on Speech Research
1993, SR-114, 81-90

The Perceptual Infrastructure of Early Phonological Development*

Alice Faber† and Catherine T. Best†

INTRODUCTION

Observation of children's vocal behavior in approximately their first two years of life reveals systematic patterns in the way they learn to speak the language spoken around them, whatever that language may be. Our purpose in this paper is to discuss some of the principles underlying this early language learning. In particular, we are interested in how and why changes take place in children's phonological inventories. We will first outline phonological development, as observed in children's babbling and early speech. Then, we will discuss a contrasting view of phonological development, based on studies of infant speech perception. Following that, we discuss some recent findings regarding the development of motor skills, also in approximately the first two years of life, and some differences between older children and adults in articulatory coordination. Finally, we will suggest that both children's limited early productive phonological inventories and the patterns of expansion of these inventories as language learning progresses do not result from increasing perceptual skill or from cognitive maturation; that is, they should not for the most part be attributed to developmental changes in linguistic rule systems. They result rather from increasing motor skill, and are, therefore, attributable to the fact that children are not just learning a language, they are also learning to talk.

Babbling to early words to full phonological inventory

The basic observation--made first in Jakobson (1941 [1968]) and reiterated by many others (see, e.g., Macken [1980] for a review)—is that children acquire the ability to produce the sounds of their native language in a lawful sequence. For present purposes, we will concentrate on the stages in (1).

(1) Canonical/reduplicative babbling
Variegated babbling
Proto-words and first words[1]
Fifty-word stage
Full phonological inventory
Adult-like phonological competence

There is (cf. Jakobson [1968]) an essential continuity in this sequence in which children learn the phonological systems and rules of their native language (Locke, 1983; Oller, 1980; Vihman et al., 1985; and, with reference to American Sign Language, Petitto, & Marentette, 1991).

In canonical and variegated babbling,[2] infants produce word-like sequences using a variety of sounds, not merely those of the ambient language. While phonotactic constraints can be observed (in particular, babbles tend to consist of one or more CV syllables), infants nonetheless make use of a relatively rich segmental inventory. However, when infants produce their first true words, around 12 months of age, their lexicons make use of a more impoverished segment[3] inventory.

We are grateful to NIH grants DC-00403 and HD-01994 for support. We would like to thank Greg Guy, Laura Ann Pettito and Joe Stemberger for helpful discussion following the oral presentation of this paper, and Andrea Levitt, Ignatius Mattingly, Michael Studdert-Kennedy, and Doug Whalen for helpful comments on an earlier version of the paper. We thank Ignatius also for the Sapir reference in footnote 13 below and Cathe Browman for discussion of the implications of using gestures as phonological primitives. All of these individuals may disagree with our interpretation of their comments. We retain, of course, full responsibility for any remaining shortcomings.

Furthermore, when their early words are compared with ambient adult models, substitutions and simplifications are evident. The phonological inventory—and the phonotactic complexity of the child's utterances—increase in parallel with lexical growth. But it is only when the infant has acquired a lexicon of approximately 50 words that minimal contrast—and thus true phonology—is likely to be observed. Children vary in how quickly they acquire the full phonological inventory of their native language. Some may do so by the age of 2 1/2, while others (of comparable intelligence) may not do so until after they have entered school. Sounds notorious for being difficult to produce are the approximants /r l y/ and fricatives, with /θ/ and /ð/ often acquired after five years of age; the contrasts among labial and anterior coronal fricatives are also late, with voicing or voicelessness preserved in substitutions (Ingram et al., 1980; Gallagher & Shriner, 1975).[4]

Accounts of phonological development

What we are interested in explaining in this paper is the constellation of facts in (2):

(2) a. Children produce rich segment inventories in babbling;
b. Children's early words are characterized by an impoverished segment inventory;
c. When children's early words are compared with their adult models, systematic patterns of substitution are observed;
d. Children's segment inventories appear to increase in terms of natural classes of segments rather than in terms of individual segments.

Even though several sorts of explanation for these facts appear in the literature, they reduce to three basic approaches, listed in (3) (similarly, Ferguson & Garnica, 1975; Strange & Broen, 1980).

(3) a. **Perception**. Children at the early stages of language do not yet accurately discriminate all of the segmental contrasts of the ambient language, and thus construct qualitatively different lexical representations from those of adults;
b. **Motor skill**. Children at the early stages of language discriminate many (or all) of the segmental contrasts of the native language, and have inferred an appropriate rule system, but lack the motor skills necessary for real time correct articulation of meaningful utterances (MacNeilage, 1980; Thelen, 1991);
c. **Rules**. Children at the early stages of language perceive many (or all) of the segmental contrasts of the native language, and have adult-like lexical representations, but they have not yet inferred appropriate phonological rule systems (similarly, Stampe, 1973, among others).

Our strategy in this paper will be as follows: We will first present evidence from studies of infant speech perception, showing that infants can, before they produce their first words, discriminate most, if not all, of the phonological contrasts of their native language. Following that, we will demonstrate that motor skill development is sufficient to explain most observed patterns of phonological inventory development. Finally, we will place our discussion in the context of current phonological models that do not and cannot rely on characteristics of linguistic rule systems to account for observed developmental patterns. We will thus argue that, in the aggregate, perceptual maturation and imperfect learning of phonological rules play a relatively minimal role in the ontogeny of mature phonological inventories. Although all of our discussion will be in terms of spoken language, we are not by any means claiming a privileged neurological or ontogenetic status for *spoken* language. Indeed, it appears that bilinguals fluent in American Sign Language and spoken English utilize similar neural substrata in both sign and speech, in contradistinction to nonlinguistic gesturing (Corina, Vaid, & Bellugi, 1992). Likewise, deaf children acquiring signed language do so in stages parallel to those in which hearing children acquire spoken language (Petitto & Marentette, 1991). We expect, therefore, that arguments parallel to ours but based on signed language would be relatively easy to construct.

PERCEPTION LEADS PRODUCTION

We will first discuss perceptual evidence that point (3) a. is incorrect; rather, prelinguistic infants are capable of detecting sound contrasts in the ambient language. One general characteristic of first language acquisition evident in the literature is that, contrary to second language acquisition, perception tends to lead production (Edwards, 1975).[5] Anecdotal evidence for this abounds (Ferguson & Garnica, 1975; Menn, 1983). In particular, a child who appears systematically to substitute /w/ for /r/, saying, for example, [wɛd] for *red*, may nevertheless recognize that an adult's target *wed* is not what he or she meant to say, and may, as a result, get annoyed that the adult fails to understand this. Such a child may, despite the apparent lack of contrast, have acoustic differences between *red* and *wed* such that the initial consonants are measurably and systematically distinct, but, nonetheless, are perceived by adults as representing the same phonemic category (Kornfeld & Goehl, 1974). In addition, Locke and Kurz (1975) find that these children often cannot distinguish their own intended *ring* and *wing*, when the tokens are randomized, and interpret this result to mean that these children are wrong in their belief that they distinguish /r/ and /w/. But, in light of Kornfeld and Goehl's findings, an alternative would be that pre-school children whom adults perceive as not distinguishing /r/ from /w/ have already acquired the adult perceptual distinction but, despite their belief that they are producing the two sounds in adult fashion, they have not yet acquired the articulatory skill necessary to production of a bunched or pharyngeal /r/ meeting adult norms.

Methods for study of infant speech perception

Study of adult speech perception involves playing sounds for subjects and asking them what they hear. This method is obviously not available for study of the speech perception abilities of prelinguistic infants. Rather it is necessary to recruit behaviors available even to very young infants, and to measure these behaviors as a reflection of the infants' time-varying interest in differences between particular classes of speech sounds. Various methods have evolved for assessing infants' interest in classes of speech sounds, and indirectly which sounds infants of different ages consider to be the same. What all of these methods have in common is that they test whether infants can hear the difference between two physically different groups of sounds. In the visual habituation paradigm, which we use for studies in our laboratory (e.g., Best, McRoberts, & Sithole, 1988; Best, 1994),[6] the infant views a brightly colored slide of a smiling person. Whenever the infant is looking at the slide, as judged by a hidden observer, sounds from one group are played over a speaker. When the infant looks away from the slide, the sounds cease, and when it looks back at the slide, the sounds return. This contingency creates a conditioned association between looking at the slide and hearing the sounds. The infant's motivation for listening to the sounds is that infants find human speech intrinsically interesting (Leavitt et al., 1973, with references). When the infant's looking falls below an individually determined threshold, that is, when it appears to have lost interest in listening to the group of sounds it has been hearing, the sounds presented are changed to the other group. Thus, if the infant has been hearing, for example, *pa...pa...pa* it might now hear *ba...ba...ba*. At this point, one of two things can happen. The infant may notice that it is hearing something new, in which case it is likely to show renewed interest in listening to the sounds, which will be reflected in increased looking at the slide. Alternatively, the infant may not notice (or care) that it is hearing a new category of sounds, in which case it will continue to show a declining interest in looking and listening. This procedure can be used with infants as young as two months of age, and, with some modification, with children, and even with adults.

The results of this procedure are interpreted as follows: Sounds that the infant discriminates potentially represent two distinct categories for the infant, and sounds that the infant appears not to discriminate may well be perceived by the infant as exemplars of a single category (assuming the infant is paying attention at all). In a typical experiment, infants in cohorts of several distinct ages will be tested. Within each age cohort, some infants will hear sounds from two conceptually distinct categories of sounds in the two phases of the experiment, and others will hear sounds that do not differ in category membership. The study of infants in several different age groups with the same experimental materials allows for the establishment of a developmental progression.

From the universal to the particular

With some systematic exceptions, infants younger than eight months old can discriminate whatever consonant contrasts they hear,[7] regardless of the phonological relevance of the contrast to the ambient language. That is, just as 6-8 month old infants being raised in an English speaking environment can discriminate between *ba* and *pa*, like English speaking adults can, so too can they discriminate between the Hindi dental and retroflex stops, a contrast that is not utilized in American English. English speaking adults cannot discriminate between the dental [t̪] and the retroflex [ṭ] (even though we can easily produce the distinction!); neither can ten month old infants.[8] But the 6-8 month old infants can, apparently with no difficulty (Werker & Tees, 1984). Janet Werker, who first observed this phenomenon, has suggested that infants younger than about 8 months of age discriminate consonant contrasts on the basis of the phonetic differences among the members of the contrast. Older infants and adults, in contrast, discriminate consonants on the basis of their *phonological* potential to distinguish lexical items in the ambient language. Between approximately 8 and 10 months of age, a perceptual reorganization takes place.[9] Thus, before a child produces its first words at approximately one year of age, the phonological structure of the language spoken around it influences the way it perceives speech sounds.

However, as we already noted, there are some contrasts, both native and non-native, that infants younger than 8 months old cannot discriminate. Four month olds cannot, for example, discriminate the native *sa-za* contrast (Eilers & Minifie, 1975; Eilers, 1977), but 6-8 month olds, tested in a different paradigm, can discriminate *sε-zε* (Best, 1994). In addition, the 6-8 month old infants cannot discriminate the Zulu fricative contrast *ɬε-ɮε* (Best, 1994). With regard to fricative place of articulation, 1-4 month olds can discriminate *sa-ša* (Eilers & Minifie, 1977), but they cannot, according to one report, discriminate *fi-θi* or *fa-θa* (Eilers, 1977);[10] Levitt et al. (1988), in contrast, found that 6-12 week olds can discriminate *fa-θa*. According to one study, 6–8 month olds cannot discriminate *fi-θi* or *fa-θa* (Eilers, 1977), although another study shows that they can form distinct categories for /f/ and /θ/, albeit less easily than they can for /s/ and /š/ (Kuhl, 1980).

In the aggregate, then, these studies show that by the time infants are starting productive use of language they can already discriminate almost all of the phonological contrasts of their native language. While they cannot yet produce adult-like forms, they appear, in many respects, to have adult-like representations, which are reflected, among other things, in their vociferous rejections of adult imitations of their phonologically impoverished productions. Nonetheless, perceptual maturation may be related to children's relatively late acquisition of fricatives, although it is not clear to us whether infants and young children have difficulty distinguishing fricative contrasts because fricatives are rare in early language, or whether fricatives are rare in early language because infants and young children have difficulty detecting fricative contrasts. In any case, the well-documented ability of pre-linguistic infants to discriminate a wide range of potentially distinctive phonetic contrasts is a crucial part of the infrastructure for their eventual development of a phonological system for their native language.[11]

WALKING PRECEDES RUNNING

We now turn to point (3) b., motor skill development. In our research, we take the position that phonological patterning is not merely a set of abstract relationships among abstract elements devoid of any essential physical characteristics. Rather, the elements in a phonological system are characterized by physiological and acoustic properties, and the relationships among them follow, at least in part, from auditory constraints on perceptual distinctiveness and neuromuscular constraints on articulator movement. In terms of perception, we observe that for a phonetic contrast to be phonologically useful it must be robust enough to be discriminated by humans using language under a variety of conditions (similarly, Thelen, 1991; Faber, 1992). While the relationship between auditory and phonetic perception is not completely understood, it is clear that they are different (Best, Morrongiello, & Robson, 1981; Repp, 1981; Mann & Liberman, 1983; Liberman & Mattingly, 1985; Werker & Logan, 1985). Furthermore, if a phonological contrast is observed in one or another language, we *infer* that it is auditorily and phonetically robust. With regard to production, we consider talking to be a motor skill comparable to walking, running, or catching a ball. And as such it must be learned. Thus, investigation of how infants and children acquire other motor skills is clearly relevant to an understanding of how they learn to talk.

Patterns of motor skill development

As our example of non-speech motor skill development, we will examine walking, following the discussion in Thelen and Ulrich (1991). Like talking, upright walking is a biologically basic, non-arbitrary human activity, that, presumably, has been selected for in the course of the evolution of our species. Skilled walking can be broken down into several component skills: i. An aggregate of muscles must be synchronously contracted; ii. The two legs must alternate between being airborne and supporting all body weight; iii. The body must maintain its balance as weight shifts between the two legs. At a finer level of detail, the alternation between the two legs in adult stepping requires a complex phasing between flexion and extension of the hips, knees, and ankles. And maintenance of balance normally requires synchronized input from the visual and vestibular systems. Newborn infants are biomechanically unsuited for walking because of their high centers of gravity and their small, weak limbs. In particular, newborn muscular activity, perhaps as a result of the constrained intrauterine environment, overwhelmingly involves muscle flexion rather than extension. However, if the needs for balance and for ankle extension are removed, by holding infants with their feet touching a backward-moving treadmill, some infants as young as one month old will stay in place by stepping forward in the alternating stepping pattern characteristic of adult walking.[12] This treadmill pattern, like the alternating kicking that young infants also engage in, is like walking; but it is not walking. It is not walking, because it does not involve all of the components of skilled walking, and because it is an involuntary response to the moving treadmill, rather than being goal-directed like skilled walking is. Walking requires an aggregate of skilled behaviors, and only when the last of these has developed does the overall skill develop, recruiting what Thelen and Ulrich (1991) refer to as skills constructed from "continuous, available precursors" (p. 44). Yet there is an essential continuity between the alternating stepping of pre-walking and of walking. What discontinuity there is results from the embedding of the alternating stepping pattern in purposeful locomotion. That is, it is a discontinuity of function rather than of movement pattern.

Speech production as motor skill

When we turn to the development of talking, that is, of skilled articulation, both continuity and discontinuity are similarly observed. The articulatory routines that children use for their early words are a subset of those that they use in babbling, so there is a continuity of motor routine. What differentiates words (and proto-words) from babbles is that the former have linguistic value and the latter do not.[13] Thus, we submit, the discontinuity between babbling and early words results from the emergence of meaning and lexicon. The sequence *dædæ*, as a word meaning *Daddy*, is embedded in a different complex control structure than as a meaningless reduplicative babble (similarly, Labov & Labov, 1978). This additional covert complexity of referential expressions *vis à vis* non-referential expressions increases the difficulty of producing what might seem to be the 'same' articulatory maneuver, and is compensated for by overt articulatory simplification.

That infants' and young children's utterances can be transcribed in terms of adult phonological categories should not mislead us into thinking that they produce these sounds in the same way adults do. Adults' transcriptions of children's speech are necessarily filtered through the adults' perceptual systems, which, in turn, are filtered by the phonological systems of the languages they speak (similarly, Macken, 1980); as a result of this modulation by adult categorical perceivers, gradual changes in children's productions may appear in transcriptions to be abrupt. As to how children are actually producing their early words, there is little relevant evidence available, due to the difficulties of interpreting acoustic analysis of utterances produced with high fundamental frequency and possibly unknown targets and of eliciting infants' cooperation in measuring articulator position or configuration during speech. Nevertheless, one recent study (Stathopoulos & Sapienza, 1991) documents differences between adults and four-year old children in respiratory and laryngeal control for speech. Perceptually, all of the children's utterances seemed normal to the experimenters; there were none of the substitutions and simplifications that characterize early phonology or the phonology of older children with language disorders. Yet the children's respiratory and laryngeal patterns for speech differed measurably and systematically from the adults'. The children, due presumably to their smaller lung volume, use a larger proportion of their vital capacity for speech breathing, and produce fewer syllables per inspiration. And the children and the adults appear to use different laryngeal settings for ordinary speech. Thus, even when children have adult phonological inventories they are not yet producing the sounds the same way adults do. While Stathopoulos and Sapienza did not study children younger than four, there is no reason to expect younger children to be more adult-like than the 4-year olds studied in their respiratory and laryngeal control. Smith (1992) likewise suggests that the greater duration of 2- and 4-year olds' utterances relative to adult controls and the greater durational variability in the children's utterances are independent reflections of the children's immature speech motor control systems.

CONCLUSION: A UNIFIED ACCOUNT OF LEARNING TO TALK

Thus far, we have argued that prelinguistic infants can discriminate most but not all of the linguistically significant contrasts in the ambient language (3 a). We have also argued that the patterns observed in beginning talkers are consistent with other patterns of motor skill development (3 b). Thus, of the potential explanations for the patterns of phonological development outlined in (2), we have eliminated perceptual learning as a general explanation, although perceptual attunement may ultimately lead to a plausible account for the late acquisition of the ability to perceive and produce contrasts among fricatives. We have also reasoned that motor skill development is a plausible explanation for the patterns of phonological development observed, and will now suggest some ways in which our account of phonological development can be related to current phonological models.

We have already noted the different control structures that it is reasonable to assume for the sequence *dædæ*, depending on whether it refers to *Daddy* or is merely a meaningless babble. This difference finds easy expression in the Articulatory Phonology of Browman and Goldstein (1986; 1989). In this view, the phonological primitives are gestures. Similar opening and closing gestures can be implemented by different articulators, and there is a difference between a complete closing and closing only to a critical position; a complete closing gives a stop, and a closing to critical position gives a fricative. One way in which this model differs from current generative models—although not necessarily so (cf. Mohanan, 1986)—is that the temporal relationship among the various gestures composing an utterance must be explicitly stated as part of the phonological representation, and these timing statements interact with non-contrastive characteristics like speech rate to bring about many of the casual speech phenomena that are, in other models, attributed to rules. Thus, in Browman and Goldstein's view, one of the things that children must learn in the course of language acquisition is the patterns of articulator phasing that are appropriate for their language. And, children's early preference for stops can be interpreted to mean that they have not yet mastered incomplete or critical closure.

For a meaningless babble *dædæ*, the tongue just happens to contact the alveolar ridge, but for the word *Daddy* a complete alveolar closure is required and produced. The physical action of the tongue tip or blade contacting the alveolar ridge might be the same in the two cases, just as the infant's prelocomotory leg kick is physically like a step; in the two cases, the pairs of actions are distinguished by the differing control regimes that they are embedded in. For a child to be able to walk, it is necessary that it be able to swing the legs in alternating fashion. Likewise, for a child to be able to produce an alveolar stop, it is necessary for it to be able to bring the front part of the tongue into contact with the alveolar ridge. However, in neither case is the second ability sufficient to guarantee the first.

In Browman and Goldstein's Articulatory Phonology, and in most versions of Generative Phonology, the basic units of phonological representation are considerably smaller than the minimal one-syllable utterance; these units are not pronounceable in isolation but only in concatenation with enough other phonological units to form a minimal utterance. In contrast, many accounts of children's phonological development suggest that children's earliest phonological representations are of larger units, Ferguson's (1986) "whole word shape" (p. 41). On this view, phonological segments of the sort generally manipulated in phonological analysis only emerge as the child's lexicon increases in size and allows for the possibility of true minimal pairs. Aside from the relatively late emergence of contrast, the primary evidence for this view is the larger scope of children's articulatory gestures, together with the different phasing relationships among these gestures (Goodell, 1991; Nittrouer, Studdert-Kennedy, & McGowan, 1989). Consequently, adults and children at the early stages of language have qualitatively different representations. We disagree. We first note the (to us) obvious point that pronunciation of even a reduplicated CVCV 'word' involves the complex sequencing of discrete and disparate articulator actions (similarly, MacNeilage & Davis, 1990). So even a global, holistic lexical representation must, in order for the child to utter it, be translated into a sequential motor program of some sort. Secondly, positing holistic representations makes it difficult to account for the well documented cases in which a child's attempts to produce a given, new word involve repeated and different permutations of some or all of the features or gestures that would be present in a full, adult representation of the word. Ferguson (1986), for example, presents 10 different attempts by a child approximately 15 months old to produce the word *pen* within a span of thirty minutes: [mãᵊ], [ˋṽ], [dɛᵈn̩], [hɪn], [ᵐbõ], [pʰɪn], [tʰn̩tʰn̩tʰn̩], [baʰ], [dʰauɴ], [buã]. While nasalization, labiality, and aspiration are variously combined in these attempts, none represents an adequate [pʰɛn]. Sequences such as this are generally taken to indicate that the child has constructed a tentative representation for *pen* containing these features, but in no particular order. It seems to us, however, that if this were the case, the child would not have made nearly so many attempts to produce these features in the sequence appropriate to more adult-like renditions of *pen*. The fact that the child made so many attempts, incorporating many of the phonological features found in its adult model, suggests that child's representation contains discrete specification of, among others, a labial closing and a velar opening, as well as the sequence in which these gestures occur, and that the child recognizes when she has implemented them in the wrong order.

Finally, the view that children at the early stages of language have qualitatively different phonological systems than the adults around them assumes a model of adult phonology that is inconsistent with modern autosegmental and metrical approaches, as described in e.g., Goldsmith (1990) (and for child phonology, Iverson and Wheeler 1987), and with Articulatory Phonology (Browman & Goldstein, 1986, 1989).

> "...the individual gestural components of articulation—the features of modern phonology—each have quite separate lives of their own, and an adequate theory of phonology will be one that recognizes this, and provides a way to understand the linkages between the individual gestures of the tongue, lips, and so forth, and larger units of organization, such as the syllable." (Goldsmith, 1990: 9)

So, in claiming an essential similarity between adults' and early language learners' phonological representations we are suggesting that neither children's nor adults' representations of words contain discrete phonological segments representable as columns in a distinctive feature matrix, of the sort posited by classical generative phonology (e.g., Chomsky & Halle, 1968). Rather, in both cases, the phonological primitives are articulator movements, or gestures, and children in the early stages of language (as in somewhat later stages) differ from adults in exactly how the gestures required for a particular word are implemented. For adult speakers, there is sufficient overlap in the gestures bringing about a particular articulatory configuration that the common idealization that speech consists of discrete segments, linearly arranged like beads on a string, does not do too much damage to the articulatory facts. For child speakers, however, the segmental idealization does more damage. But the difference between children and adults resides in the amount of gestural overlap (Nittrouer, Studdert-Kennedy, and McGowan, 1989; Goodell, 1991), not, we would claim, in the nature of the phonological representation that is most appropriate for each.

Despite the overwhelming evidence in favor of a motor-skill-based account of phonological acquisition, rule-based accounts of differences between children's and adults' phonological systems must still be considered (3 c.). The attractiveness of such accounts, as noted by Stampe (1973), follows from the lawful nature of the relationship between mature and immature phonological systems. And, indeed, the systematic nature of this relationship underlay the development of Stampe's natural phonology, according to which language acquisition consists in large measure in suppressing innate natural processes. On this account, the primary difference between mature and immature phonologies is that children with small phonological inventories have not yet learned to suppress those phonological processes that do not apply in their native language.

Menn (1983) takes a somewhat different approach. Essentially, her proposal is that children at the early stages of language are subject to severe output constraints on possible phonological forms. These output constraints are presumably similar to Surface Phonetic Constraints (MSCs) of the sort proposed by Shibitani (1973) to account for phonotactic regularities. Such MSCs are at least partially language specific (some languages, for example, allow word initial consonant clusters, and some do not), and, hence, must be considered part of the grammar of a language. Although it may appear that Menn is thus claiming that children's lexical representations are different from adults', she is not. These output constraints restrict what children can say in various ways, including, in addition to modification and deletion of segments, avoidance of lexical items that a child cannot yet pronounce. While the output constraints for a particular child at a particular point in time are clearly not universal, Menn nonetheless sees them as outside the child's developing grammar; rather, they in some sense represent a metalinguistic codification of the child's articulatory capacities.

Either Stampe's or Menn's approach may appropriately capture the range of regularities in early phonology and in children's attempts to produce understandable words; however, it is not clear that they vitiate the need to appeal to motor skill development to account for some acquisitional patterns. Furthermore, many of the surface differences that lead Stampe and others (e.g., Stemberger, 1988; Matthei, 1989) to posit different rule systems for children and adults may simply reflect children's gestures being implemented with different phasing relations than adults' gestures are. We suppose, in addition, that this formal difference reflects underlying motor skill differences between children and adults, rather than constituting *per se* a crucial difference between children's and adults' linguistic skills. Our account of increasing segment inventories in early language, then, relies on the difficulty inherent in embedding previously mastered articulatory maneuvers in the new, hierarchical control structure of non-linear phonology or of a gestural score. In terms of generative phonology, an alternative could be proposed utilizing underspecification (Archangeli, 1988, and, with regard to child language, Iverson and Wheeler, 1987, Stemberger & Stoel-Gammon, 1991). That is, children's lexical representations contain less phonological information than do adults' and therefore, more putatively redundant information is specified by rule. Thus, all vowels could, for example, be specified as dorsal, and the feature [low] provided by rule, not merely in cases in which adults have a low vowel, but across the board. Likewise, consonants could be unspecified for place and manner, and the feature labial would be specified by rule in all cases and continuant would be specified in no cases.[14] Such an alternative is, we believe, incorrect, in that it supposes that children's representations differ systematically from adults' and in that it assumes that children who cannot produce the contrast between *Daddy* and *doggy* cannot perceive it. That the latter assumption is clearly wrong suggests that the former is as well.

We close with one final point. Our suggestion in this paper has been that the pervasive and systematic phonetic differences between children's early linguistic forms and those of adults, as well as the developmental path by which children finally arrive at the normative forms of the ambient language primarily reflect developmental differences in motor skill, in particular in articulatory agility. To the extent that these differences can be formalized in current models of phonology, one is tempted to attribute them to cognitive rather than motor skill immaturity. We have argued that this would be mistaken. Despite this, we are unwilling to take the further step of claiming that no developmental phonological phenomena can be attributed to infants' immature perceptual systems or to their construction of inappropriate or overgeneralized rules. Indeed, the latter phenomena are well-documented in morphology and syntax, although perhaps not as pervasive as is generally thought (see Marcus et al. [1992] for details). We would like to suggest, however, that only when those aspects of phonological development that result from motor skill development are factored out will it be possible properly to understand the true roles of perceptual maturation and grammar construction in phonological development.

REFERENCES

Abbs, M. S., & Minifie, F. D. (1969). Effects of Acoustic Cues in Fricatives on Perceptual Confusions in Preschool Children. *Journal of the Acoustical Society of America, 46*, 1535–1542.

Archangeli, D. (1988). Aspects of underspecification theory. *Phonology*, 5, 183–207.

Best, C. T. (1994). Emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. Goodman & H. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167-224). Cambridge, MA: MIT Press.

Best, C. T., Morrongiello, B. A., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and non-speech perception. *Perception & Psychophysics, 29*, 191–211.

Best, C. T., McRoberts, G. W., & Sithole, N. (1988). The phonological basis of perceptual loss for non-native contrasts: Maintenance of discrimination among Zulu clicks by English speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 345–360.

Browman, C. P., & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology Yearbook, 3*, 219–252.

Browman, C. P., & Goldstein, L. M. (1989). Articulatory gestures as phonological units. *Phonology, 6*, 201–251.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English.* New York: Harper and Row.

Corina, D. P., Vaid, J., & Bellugi, U. (1992). The linguistic basis of left hemisphere specialization. *Science, 255*, 1258–1260.

Edwards, M. L. (1975. Perception and production in child phonology. *Journal of Child Language, 1*, 205–219.

Eilers, R. E. (1977). Context-sensitive perception of naturally produced stop and fricative consonants by infants. *Journal of the Acoustical Society of America, 61*, 1321–1336.

Eilers, R. (1980). Infant speech perception: History and mystery. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child Phonology*, Vol. 2: *Perception*, (pp. 23–39). New York: Academic Press.

Eilers, R. E., & Minifie, F. D. (1975). Fricative discrimination in early infancy. *Journal of Speech and Hearing Research, 18*, 158–167.

Eilers, R. E., & Oller, D. K. (1976). The role of speech discrimination in developmental sound substitutions. *Journal of Child Language*, 3, 319–329.

Eimas, P. D., Siqueland, E. R. Jusczyk, P., & Vigorito, J. (1971). Speech Perception in Infants. *Science, 171*, 303–306.

Faber, A. (1992). Articulatory variability, categorical perception, and the inevitability of sound change. In G. Davis, & G. Iverson (Eds.), *Explanation in historical linguistics* (pp. 59-75). Amsterdam: John Benjamins.

Ferguson, C. A. (1986). Discovering sound units and constructing sound systems: It's child's play. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 36–53). Hillsdale, NJ: Erlbaum.

Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language 51*, 419–439.

Ferguson, C. A., & Garnica, O. K. (1975). Theories of phonological development. In E. H. Lenneberg & E. Lenneberg (Eds.),

BEST COPY AVAILABLE

Foundations of language development (pp. 153–180). New York: Academic Press.

Gallagher, T. M., & Shriner, Thomas H. (1975). Articulatory inconsistencies in the speech of normal children. *Journal of Speech and Hearing Research, 18,* 168–175.

Goldsmith, J. A. (1990). *Autosegmental and metrical phonology.* Chicago: University of Chicago Press.

Goodell, E. W. (1991). *Gestural organization in the speech of 22- 32-month old children.* Unpublished doctoral dissertation, University of Connecticut.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologica, 9,* 317–323.

Ingram, D., Christensen, L., Veach, S., & Webster, B. (1980). The acquisition of word initial fricatives and affricates in English children between 2 and 6 years. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child Phonology, Vol. 1: Production* (pp. 169–192). New York: Academic Press.

Iverson, G., & Wheeler, D. (1987). Hierarchical structures in child phonology. *Lingua, 73,* 243–257.

Jakobson, R. 1941 [1968]. Child language, aphasia and phonological universals. Tr. by Allan R. Keiler. The Hague: Mouton.

Kornfeld, J. R., & Goehl, H. (1974). A new twist to an old observation: Kids know more than they say. Chicago Linguistic Society Parasession on Natural Phonology (pp. 210–219).

Kuhl, P. K. (1980). Perceptual constancy for speech-sound categories in early infancy. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A.,Ferguson (Eds.), *Child phonology,* Vol. 1: *Perception* (pp. 41–66). New York: Academic Press.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255,* 606–608.

Labov, W., & Labov, T. (1978). The phonetics of *Cat* and *Mama. Language, 54,* 816–852.

Leavitt, L. A., Morse, P. A., Brown, J. W., & Graham, F. K. (1973). Cardiac orienting to speech and non-speech stimuli in 6-week-old infants. *Pediatric Research, 7,* 419.

Levitt, A., Jusczyk, P. W., Murray, J., & Carden, G. (1988). Context effects in two-month-old infants' perception of labiodental/interdental fricative contrasts. *Journal of Experimental Psychology: Human Perception and Performance,* 14, 361–368.

Liberman, A. M., & Mattingly, I. G. (1985). The Motor Theory of Speech Perception revised. *Cognition, 21,* 1–36.

Locke, J. L. (1978). Selective loss of phonetic production and perception: An index to the child's acquisition of phonology. *Journal of the National Student Speech and Hearing Association,* 3–11.

Locke, J. L. (1983). *Phonological acquisition and change.* New York: Academic Press.

Locke, J. L., & Kurtz, K. J. (1975). Memory for speech and speech for memory. *Journal of Speech and Hearing Research, 18,* 176–191.

Macken, M. A. (1980). Aspects of the acquisition of stop systems: A cross-linguistic perspective. In G. H. Yeni-Komshian, J. F. Kavanagh, , & C. A. Ferguson, (Eds.), *Child Phonology,* Vol. 1: *Production,* (pp. 143–168). New York: Academic Press.

MacNeilage, P. F. (1980). The control of speech production. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson, (Eds.), *Child Phonology,* Vol. 1: *Production* (pp. 9–21). New York: Academic Press.

MacNeilage, P. F., & Davis, B. (1990). Acquisition of speech production: Frames, then content. In M. Jeannerod (Ed.), *Attention and performance XIII: Motor representation and control* (pp. 453-476). Hillsdale, NJ: Erlbaum.

Macken, M. A. (1980). Aspects of the acquisition of stop systems: A cross-linguistic perspective. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology,* Vol. 1: *Production,* (pp. 93–112). New York: Academic Press.

Mann, V. A., & Liberman, A. M. (1983). Some differences between phonetic and auditory modes of perception. *Cognition, 14,* 211–235.

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., & Xu, F. (1992). *Overgeneralization in language acquisition.* Society for Research in Child Development, Monograph #228.

Matthei, E. H. (1989). Crossing boundaries: More evidence for phonological constraints in early multi-word utterances. *Journal of Child Language, 16,* 41–54.

Menn, L. (1983). Development of articulatory, phonetic, and phonological capabilities. In B. Butterworth (Ed.), *Language production,* Vol. 2: *Development, writing and other language processes* (pp. 3–50). New York: Academic Press.

Mohanan, K. P. (1986). *The theory of lexical phonology.* Dordrecht: Reidel.

Nittrouer, S., Studdert-Kennedy, M., & McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research, 32,* 120–132.

Oller, D. K. (1980). The emergence of the sounds of speech in infancy. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology,* Vol. 1: *Production,* 93–112. New York: Academic Press.

Petitto, L. A. & Marentette, P. F. (1991). Babbling in the manual mode: Evidence for the ontogeny of language. *Science, 251,* 1493–1496.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 421-435.

Repp, B. H. (1981). Two strategies in fricative discrimination. *Perception & Psychophysics, 30,* 217–227.

Sapir, E. (1925). Sound patterns in language. *Language, 1,* 37–51.

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English. *Applied Psycholinguistics, 3,* 243–261.

Shibitani, M. (1973). The role of surface phonetic constraints in generative phonology. *Language, 49,* 87–106.

Smith, B. L. (1992). Relationships between duration and temporability in children's speech. *Journal of the Acoustical Society of America, 91,* 2165–2174.

Stampe, D. (1973). *A dissertation on natural phonology.* Unpublished University of Chicago Ph.D. Dissertation.

Stathopoulos, E., & Sapienza, C. 1991. Respiratory and laryngeal function of children and adults during vocal intensity variation. Presented at the Annual Meeting of the American Speech-Language and Hearing Association, Atlanta.

Stemberger, J. P. (1988). Between-word processes in child phonology. *Journal of Child Language, 15,* 39–61.

Stemberger, J. P., & Stoel-Gammon, C. (1991). The underspecification of coronals: Evidence from language acquisition and performance errors. In C. Paradis & J.-F. Prunet (Eds.), *Phonetics and phonology 2: The special status of coronals* (pp. 181-199). San Diego: Academic Press.

Strange, W., & Broen, P. A. (1980). Perception and production of approximant consonants by 3-year olds. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology,* Vol. 2: *Perception* (pp. 117–154). New York: Academic Press.

Thelen, E. (1991). Motor aspects of emergent speech: A dynamic approach. In N. A. Krasnegor, D. M. Rumbaugh, R. L. Schiefelbusch, & M. Studdert-Kennedy (Eds.), *Biological and*

BEST COPY AVAILABLE

behavioral determinants of language development (pp. 339–362). Hillsdale, NJ: Erlbaum.

Thelen, E., & Ulrich, B. D. (1991). *Hidden skills: A dynamic systems analysis of treadmill steps during the first year*. (Society for Research in Child Development, Monograph 223).

Vihman, M. M., Macken, M. A., Miller, R., Simmons, H., & Miller, J. (1985). From babbling to speech: A reassessment of the continuity issue. *Language, 61*, 397–445.

Werker, J. F., & Logan, J. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, 37, 35–44.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*, 49–63.

FOOTNOTES

[*] In R. Corrigan, G. Iverson, & S. D. Lima (Eds.), *The reality of linguistics rule* (1994).

[†] Also Wesleyan University.

[1] Proto-words function like words, but differ from them in having no obvious adult prototypes. See Menn (1983) for details.

[2] Canonical babbling is characterized by a relative lack of within-utterance variation in consonants and vowels (Oller, 1980). Canonical babbles are likely to be reduplicated utterances like *bæbæbæbæbæ*, while variegated babbles have more varied phonological structure.

[3] Here and throughout, our use of the term *segment* as an expository shorthand in no way involves an explicit claim that the child's early words are constructed bottom-up from discrete phonological units rather than being holistic units which are only metalinguistically analyzable into component segments.

[4] The ability to perceive fricative contrasts is also acquired relatively late. For example, 12-14 month old infants can discriminate *fi-θi* but not *fa-θa* (Eilers, 1977). Two year olds cannot discriminate *fɪš-θɪš* (Eilers & Oller, 1976) and 3-5 year olds (Abbs & Minifie, 1969) have difficulty with /f/-/θ/ and /v/-/ð/.

[5] For production abilities leading perception abilities in second language learning, see Goto (1971) and Sheldon and Strange (1982).

[6] Other common procedures involve the non-nutritive sucking paradigm (Eimas et. al., 1971) and the Visually Reinforced Head-Turn paradigm (Kuhl et al., 1992). In addition, some researchers monitor attention-related changes in heart rate. See Eilers (1980) for further details on these paradigms. The conditioned visual fixation paradigm described in the text is the most flexible, in that it can be used with infants in a wide variety of age groups, simplifying longitudinal comparison.

[7] Less attention has been paid to the perception of non-native vowel contrasts. Kuhl et al. (1992) concerns the development of prototypes for native (but not, of course, for non-native) vowel phonemes. We are currently doing pilot work for a study of infants' discrimination of non-native vowel contrasts, and a similar study has recently been completed by Polka and Werker (1994).

[8] Likewise, 3-year old children can discriminate neither [t̪] from [t] nor [ʈ] from [t] (Locke, 1978).

[9] Werker's conclusions are with reference to consonantal phonology. The results of Kuhl et al. (1992) suggest that native language effects on vowel perception may be observable in infants as young as six months of age.

[10] These same 5–16 week old infants could discriminate [as]-[a:z], in which a naturalistic vowel length distinction supplements the fricative voicing contrast; the [a:s]-[a:z] data from comparable infants show a non-significant tendency toward discrimination (Eilers, 1977). Levitt et al. (1988) suggest that the failure of Eilers' young subjects to discriminate *fa-θa* might be the result of degraded stimuli; only c. 70% of the adults tested could correctly identify the stimuli.

[11] While our position is superficially similar to that of Eimas (e.g., Eimas et al. 1971), we do not mean to claim that phonetic categories are innate, but rather that the cognitive ability to discern categories in the environment is.

[12] In Thelen and Ulrich's (1991) study, only some infants exhibited alternating leg movements at one month of age; by seven months, however, all infants in the study did, at least at some treadmill speeds.

[13] This distinction is comparable to Sapir's (1925) distinction between producing a voiceless bilabial approximant [hw] and blowing out a candle. While the action may in some sense be the same in the two cases, only in the former is approximating the two lips required by a particular communicative intent.

[14] The sequence in the text is one of many possible within underspecification theories; one child might at first implement all stops as labial, and another as coronal. Nothing in the theory implies, to our knowledge, a particular order of acquisition.

Haskins Laboratories Status Report on Speech Research
1993, SR-114, 91-100

The Stability of Distinctive Vowel Length in Thai*

Arthur S. Abramson†

INTRODUCTION

Many languages have phonological distinctions of quantity in consonants or vowels or both. Among them, Italian is known for its word-medial intervocalic short and long consonants, while Pattani Malay (Abramson, 1987) has word-initial prevocalic short and long consonants. Swedish, some dialects of German, and Thai have short and long vowels. Finnish has a length distinction for both consonants and vowels. Such distinctive length in segments is to be distinguished, of course, from other communicatively relevant roles of timing in speech, e.g., in stress and intonation.

The obvious physical correlate of the length distinction in phonetic segments is relative duration. That is, in the simplest case, the articulatory configuration is held longer for the "long" segment than for the "short" one. Limiting our attention here to vowels, we note an important observation made by Daniel Jones (1950, p. 28): "In languages where vowel length is significant it very often happens that the quality of a long vowel is not quite the same as that of the corresponding short vowel." Ilse Lehiste (1970, pp. 30–33) amplifies the point by commenting that in "quantity" languages some differences in the phonetic quality of short and long vowels can he observed, although such languages differ somewhat in the amount of correlation between length and quality. To the extent that relative duration *is* the primary differentiator of the two classes of vowels, some linguists may prefer to handle the timing difference phonologically as one of gemination rather than distinctive length. Gemination means that what I have been calling a long segment is in fact a sequence of two instances of the same speech sound. This implies rearticulation at the onset of the second occurrence of the segment. Auditory impressions and acoustic observations suggest strongly that such rearticulation is highly unlikely; nevertheless, whether or not such an argument is tenable phonetically is not a likely outcome of the data to be presented in this paper.[1]

The language of concern here is Standard Thai, the official language of Thailand. It is the standard variety of Central Thai, the regional dialect of Bangkok and a sizable area around it. Traditional Thai grammar posits nine short vowels and nine long counterparts, as well as various diphthongs and vowel clusters. Linguists working on the language, both Thai and foreign, generally accept this view, although some may prefer to transcribe the long vowels as geminates (Tingsabadh & Abramson, in press).

In my own early experimental phonetic approach to Thai (Abramson, 1962; cf. also Abramson, 1974), I examined the vowel-length contrast in isolated vowels, word-pairs in carrier sentences, and a small sampling of running speech. The resulting acoustic data clearly supported relative duration as the major differentiator of the two classes of vowels. The average ratio of long vowels to short vowels was 2.9 for isolated vowels, 2.5 for the pairs in carriers, and 2.5 for running speech. In addition, experiments in perception demonstrated that for native speakers of the language relative duration provides a sufficient auditory "cue" for this phonemic distinction. At that time, the stimuli for the listening tests were made by shortening original long vowels in minimal pairs of words to values within the

---

This work was supported by Grant HD-01994 from the National Institute of Child Health and Human Development to Haskins Laboratories, New Haven, Connecticut. I thank my friends and colleagues in the Faculty of Arts, especially the Department of Linguistics, of Chulalongkorn University, Bangkok for all their help in this project. I had advice from Professor Leonard Katz in the statistical analysis, as well as help from Dr. Betty Kollia in improving the graphic displays and from Ms. Yvonne Manning-Jones in the preparation of the manuscript.

ranges of their short counterparts. More recently (Abramson & Ren, 1990), computer-manipulation allowed us also to lengthen original short vowels incrementally. Work by other investigators (Sittachit, 1972; Saravari & Imai, 1983; Gandour, 1984; Gandour & Dardarananda, 1984; Gandour, Weinberg, Petty, & Dardarananda, 1987; Svastikula, 1986) confirms the role of relative duration.

THE ROBUSTNESS OF ACOUSTIC CUES

The work being presented here is part of a larger endeavor, one that seeks to investigate the stability of acoustic cues to phonemic distinctions in a range of styles of speech. The term *acoustic cue* or just *cue* was coined by the Haskins Laboratories group in the early fifties.[2] Acoustic analysis of utterances in a language should yield certain properties that differentiate one class of phonemes from all other classes in the system; furthermore, a more detailed breakdown of each such class should reveal subcategories of such properties that serve to differentiate the phonemes within the class. Experiments may show that these properties not only separate phonemes in speech production but are also sufficient to distinguish them in perception. The latter does not automatically follow from the former, since a phonemic distinction could rest on several properties with varying amounts of power as information-bearing elements for perception. A property with such power in speech perception is called an acoustic cue. Examples are shifts upward and downward in frequency of formants (resonances of the vocal tract) for the place of articulation of stop consonants, relative frequency-heights of formants for vowels, spectral location—higher or lower in frequency—and extent of frication energy for fricatives, and, for our purposes here, the relative durations of vocalic stretches for the contrast between short and long vowels.

To this day, most of what we know about the acoustic properties of speech signals and their value as cues, as well as the underlying motor behavior controlled by various physiological mechanisms, comes from the study of short utterances carefully recorded in the laboratory. Such utterances are likely to be isolated words, short expressions, or key words embedded in a carrier sentence. For perception testing, such utterances may be manipulated on the computer along certain dimensions, although most experimental work on perception has used synthetic speech. In perceptual experiments, the listeners' choice of responses may be words or even nonsense syllables that are phonologically "legal" within the language.

In some kinds of phonetic research, for example, prosody, it has long been recognized that one must work with longer spans, usually sentences but maybe even a whole discourse. Much less has been done, however, in the study of vowels and consonants in running speech or even in other styles that are not citation forms. One expected characteristic of spontaneous speech is less articulatory precision than in citation forms; nevertheless, in the very same spontaneous style of speech a need, from time to time, to be very clear or emphatic may yield somewhat greater precision in the control of articulatory dynamics than in ordinary citation forms. In addition, in unrehearsed running speech, whether casual or deliberate, there is much top-down information from the phonological, morphological, syntactic, and pragmatic contexts. In the classical experiments on the cues, most of the top-down information was kept out of play through the use of isolated citation forms.[3] The work presented here is part of an effort to pursue implications in the literature (Barik, 1977; Levin, Schaffer, & Snow, 1982; Remez, Berns, Nutter, *et al.*, 1991; Laan, 1992) that acoustic differences between spontaneous and read speech are complex. The plan is to study how well phonemic distinctions, as they have been analyzed in citation-form speech in the past, are preserved phonetically in running speech. Furthermore, for the many phonemic distinctions that are no doubt well maintained, we ask whether the acoustic properties linked with the distinctions are easily derived from the cues found in traditional speech-perception research.

The foregoing matters are complicated by overlap between styles. Thus, speech read from written material includes both citation forms and the more or less fluent reading aloud of texts. (Of course, skilled actors can make read or memorized speech sound quite spontaneous.) Running speech includes both read speech and spontaneous talking. Somewhere between the last two is to be fitted the giving of a formal lecture not from a written text but from an outline. Speakers apparently vary widely in the care with which they project bottom-up phonetic information across these styles. The phonetic precision and thus, perhaps, the perceptibility, of a word is often correlated with recent occurrence of the word in the discourse, familiarity of the topic to the listener, complexity of a task to be performed, surrounding noise level, and other such factors (Lieberman, 1963; Barik, 1977; Levin, Schaffer &

Snow, 1982; Fowler & Housum, 1987; Fowler, 1988; Anderson, Bader, Bard, *et al.*, 1991; Remez *et al.*, 1991; Laan, 1992; Kohler, 1992).

My attention will be restricted here to the acoustic examination of the robustness of relative duration as a differentiator of phonemically short and long vowels in Thai. Inasmuch as vowels are notoriously vulnerable to expansion and compression in time as speakers vary their rates of articulation, their speaking styles, their focus on different parts of the discourse, the extent to which a vowel-length distinction is maintained through the features of relative duration alone ought surely to be, in its simplicity, an excellent starting point for my investigation of the robustness of acoustic cues. Other factors, such as formant patterns, that might also serve as cues, even if secondary ones, to a vowel-length distinction (e.g., Straka, 1959; Hadding-Koch & Abramson, 1964; Bennett, 1968; Abramson & Ren, 1990) will not be treated here. Words embedded in short carrier sentences, short expressions, and spontaneous casual conversation will be examined. Although the data should have implications for perception, experiments testing perceptual hypotheses derived from the findings are planned for a sequel to the present study. These hypotheses could include the relevance of other phonetic characteristics in addition to duration.

Procedure

Eight pairs of Thai words, each pair minimally distinguished by vowel length, were recorded in semantically appropriate carrier sentences by four educated native speakers of Central Thai. The words and a sampling of the sentences are shown in Table 1.[4] The sentences were recorded in a random order. For the first reading, the speakers were asked to use a normal, comfortable rate. For the second reading, they were asked to read faster. Each list of sentences was recorded twice by each speaker. Although in such a procedure the speaking rates were likely to differ widely from speaker to speaker, it was felt that self-determination of normal and fast rates would make for more natural productions.

Table 1. *Minimal pairs of words in sentences.*

Words

cìp	'to sip'	cì:p	'to flirt'
hèt	'mushroom'	hè:t	'cause'
tàk	'to dip up'	tà:k	'to dry'
cām	'to remember'	cā:m	'to sneeze'
khǎj	'to unlock'	khǎ:j	'to sell'
khùt	'to dig'	khù:t	'to scrape'
thūn	'fund'	thū:n	'to carry on the head'
sòt	'fresh'	sò:t	'unmarried'

Sample of Sentences

phájājā:m hǎ: hèt hâj khūn	'I'm trying to find mushrooms for you.'
phájājā:m hǎ: hè:t hâj khūn	'I'm trying to find reasons for you.'
jà: khùt mâ:k kɤ̄:n pāj	'Don't dig too much.'
jà: khù:t mâ:k kɤ̄:n pāj	'Don't scrape too much.'
mâj sâ:p sòt rɯ́ plà:w	'I don't know whether it's fresh.'
mâj sâ:p sò:t rɯ́ plà:w	'I don't know whether he's single.'

To obtain enough unrehearsed conversational speech, I found four members of the staff of Chulalongkorn University, two women and two men, who knew each other well, were quite used to microphones, and did not mind chatting informally about things of interest to them. Two at a time, a man and a woman, sat in a recording booth and talked to each other for ten to fifteen minutes about such topics as events on campus, plans for projects, and vacations. Their speech seemed very natural, varying widely in tempo, emphasis, and clarity. Some of it, not surprisingly, was unusable because of overlapping utterances, laughter, and other distortions.

With the help of a Thai colleague, one of my four speakers, I went through the recorded conversations and wrote down a number of words and short expressions uttered by each person. Then, one by one, I had each person read his or her excerpts into a tape recorder. Although this material included phrases and short sentences, it is probably best viewed as a set of citation forms.

Unfortunately, the literature does not reveal a universally accepted criterion for the measurement of vowel duration in spectrograms or waveforms. One common practice is to measure only that span of the vocalic formant pattern that is voiced, i.e., excited by glottal pulsing. Such a definition makes a partly or wholly unvoiced vowel impossible. Thus, for example, there would be no vowels in whispered speech! Others, rejecting that definition, measure the time during which the supraglottal vocal tract appears to maintain a relatively open configuration, one without local constriction or closure of the kind that yields consonants, no matter what the source of acoustic excitation is. Thus, working with the latter articulatory bias, I have measured every vowel from the release of the prevocalic consonant to the end of the formant pattern. If the syllable ends in a consonant, the sudden ending of the formants, perhaps with a visible upward or downward transition of one or more of the formants, signals the moment of closure. When necessary, help can be had by comparing the spectrogram with a waveform of the utterance.

In Figure 1 we see two Thai words taken from their carrier sentences. Both begin with a voiceless aspirated dorso-velar stop. The major difference between this and its voiceless *unaspirated* counterpart is that the latter would show voicing onset immediately after the release instead of the turbulence seen in the spectrograms and waveforms of Figure 1 (Abramson, 1989).

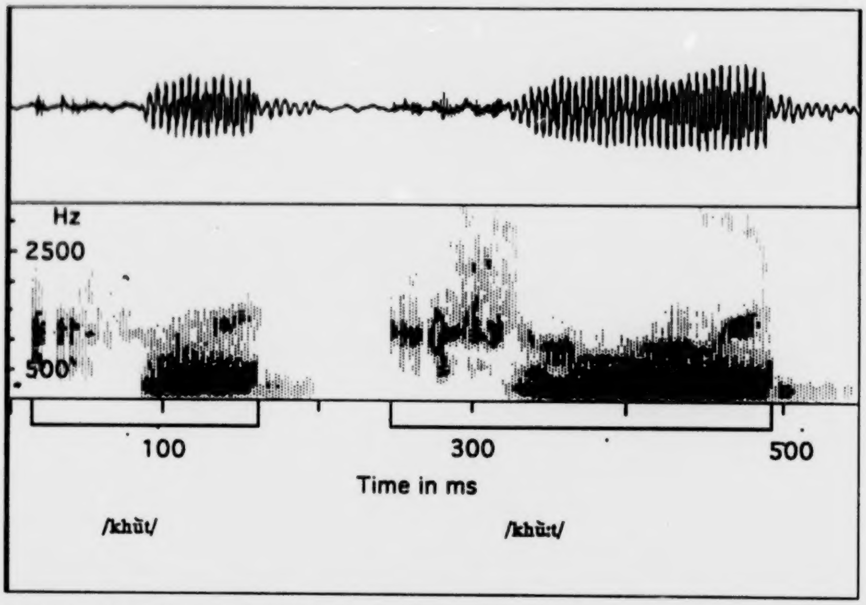


Figure 1. **A minimal pair of words cut from their carrier sentences. Waveforms above and spectrograms below.**

Thus, for the examples in the figure, as well as for voiced and voiceless unaspirated initials, the vowel onset is taken to be the release of the initial consonant. In the figure the second formant in each word moves upward for the final dental closure. The rectangles under the spectrograms show the vocalic spans, which include the aspirated voicing lags determined by voice onset time (VOT) (Lisker & Abramson, 1964). As a statistical test of the validity of this approach, I have measured the VOTs of all the aspirated voiceless stops of the four speakers who recorded the minimal pairs of words in the carrier sentences. I limited the test to the normal speech rate. This balanced set of words with short and long vowels yielded 24 tokens of each length. One might argue that my criterion for determining the duration of a vowel would be undermined by a finding of significantly larger VOT values for short vowels, since this would make the two length categories less different. In fact, the opposite tends to be true. The short vowels had a mean VOT of 59 ms, while the long vowels had a mean VOT of 66 ms, with considerable overlap of the standard deviations. A paired *t*-test showed the difference to be only marginally significant (t (23) = -1.99, $p < 0.06$).

A special problem arises in the handling of diphthongs. In a diphthong we have a gliding articulatory movement to or from a vowel target. If the vocal-tract shape of the target is held for a bit, it will be reflected in essentially steady-state formants. Only when such steady states are available do I measure the duration of the target vowel of a diphthong. Many such words with movement throughout the vocalic portion had to be left unmeasured in the running speech; any estimate of a segmentation point was simply too unreliable to inspire confidence. Indeed, this may be seen as an example of the caution that is needed in undertaking the task of chopping a speech signal into spans that are said to correspond to phonetic segments.

Results

The data for the eight minimal pairs of short and long vowels in words in carrier sentences are given graphically in Figure 2. The means and standard deviations of the measurements are given for both the normal ("slow") and fast rates of speech. The ratios of long to short vowels are 1.8 for the slow rate and 1.5 for the fast rate. The data were put through an analysis of variance. Rate as a factor was significant, $F(1,3) = 28.5$, $p < 0.02$. That is, the fast short and long vowels were both significantly shorter than their slow counterparts. Vowel length is also significant, $F(1,3) = 568.7$, $p < 0.001$. The interaction of rate and length is significant, $F(1,3) = 49.9$, $p < 0.006$. The identity of a word in the set of 16 words was significant, $F(7,21) = 45.4$, $p < 0.001$, but not the identity of a particular token of a word, $F(1,3) = 0.07$, $p = 0.8$, n.s. There is also a significant interaction between word and length, $F(7,21) = 3.6$, $p < 0.02$. The means and standard deviations of the vowel durations measured in the conversational excerpts that were separately recorded by all four speakers are given in Figure 3. The ratio of long to short vowels is 2.2. The results were shown by unpaired *t*-tests to be highly significant for each of the speakers, as seen in Table 2.

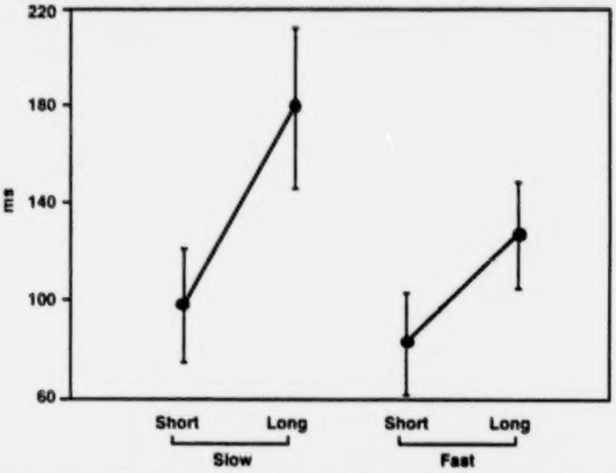


Figure 2. Means (dots) and standard deviations (vertical bars) of the durations of eight minimal pairs of long and short Central Thai vowels uttered in carrier sentences at two rates, normal ("slow") and fast by four speakers. N=64 for each point.

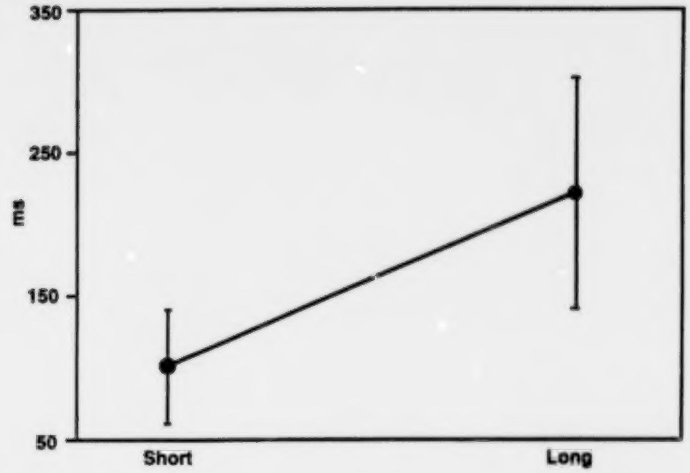


Figure 3. **Means and standard deviations of 82 short and 67 long vowels in words and phrases read by four speakers.**

Table 2. *Means, Standard Deviations, and Significance Levels for the Words and Phrases Read by the Four Speakers: Unpaired t-Tests.*

	Number		Short		Long				
Spkr	Short	Long	M	SD	M	SD	*df*	*t*	*p*
TL	27	23	101	41.8	222	52.1	48	-9.2	<.001
PM	9	18	111	45.8	233	108.0	25	-3.2	<.004
SA	11	11	108	44.2	205	66.6	20	-4.1	<.001
ST	35	15	98	36.3	224	97.1	48	-6.7	<.001

As for the running speech, the means and standard deviations of the vowel durations are given for all four speakers in Figure 4. The ratio of long to short vowels for the data in the figure is 2.1. The results were shown by unpaired *t*-tests to be highly significant for each of the speakers, as can be seen in Table 3.

It is necessary now to digress for a moment and state that these data have been taken only from vowels that I could identify with great confidence as short or long. I hasten to add that what should seem obvious is not so simple a matter in Thai running speech. That is, we cannot always decide on the basis of its citation form or dictionary entry which length the vowel of a morpheme has in an utterance. There are rule-governed shifts from long vowel to short in non-final morphemes in compound nouns and reduplicated adverbials (Sutadarat, 1978, pp. 70–71). In addition, in very casual speech there is a tendency to have weak stress with shortening of lexical long vowels in other constructions and in certain syntactic classes of words that are unstressed, such as particles, negatives, and some adverbs (Sutadarat, 1976, pp. 149–150). Contrariwise, some of the latter that are lexically short may become long under emphatic stress. Where such processes are evident, I have not hesitated to assign the resulting vowels to the deviant category. For all others, I have assigned them to the category found in the lexical entry or citation form, even when a "long" vowel seemed surprisingly short for its context, or a "short" vowel surprisingly long. Such a criterion, it seems to me, must be adopted if one is to run a fair test of the stability of the length distinction. One good outcome of this study would be an attempt by phoneticians and phonologists, especially those who are native speakers of Thai, to formulate stringent criteria for handling the matter. In the meantime, I believe that the dubious cases are few enough not to affect the results seriously.

The ratio of T.L.'s long to short vowels in running speech, i.e., for her data in Figure 4, is 1.9. Since she is the only speaker whose vowels have been measured under all three conditions, it may be of some interest to compare her data between conditions. This was done by means of unpaired *t*-tests.

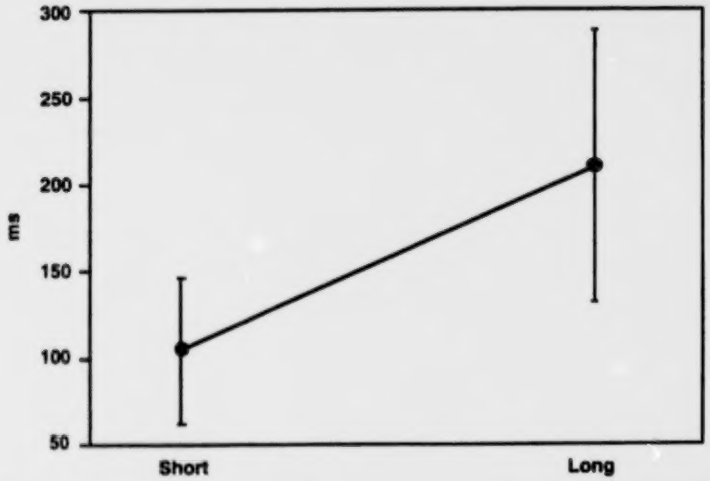


Figure 4. **Means and standard deviations of 438 short and 381 long vowels in the running speech of all four speakers.**

Table 3. *Means, Standard Deviations, and Significance Levels for the Running Speech of the Four Speakers: Unpaired* t*-Tests.*

	Number		Short		Long				
Spkr	Short	Long	M	SD	M	SD	*df*	*t*	*p*
TL	146	133	110	49.4	213	92.7	277	-11.8	<.001
PM	55	46	96	35.0	206	64.1	99	-11.0	<.001
SA	81	77	91	27.8	192	60.6	156	-13.5	<.001
ST	156	125	113	38.9	219	77.0	279	-15.1	<.001

First, let us compare the read excerpts and the slow and fast minimal pairs in sentences. In the comparison with the slow pairs, the difference is not significant for the short vowels (t (41) = -0.3, p = 0.7, n.s.), but it is significant for the long vowels (t (37) = 2.6, $p < 0.02$). For the fast pairs, again the difference between the two sets of short vowels is not significant (t (40) = 1.3, p = 0.21), while for the long vowels it is highly significant (t (37) = 6.3, $p <$ 0.001).

Comparisons of the excerpts and the minimal pairs with running speech yield mixed results. There is no significant difference between the excerpts and running speech either for the short vowels (t (171) = -0.9, p = 0.37, n.s.) or for the long vowels (t (154) = 0.5, p =0.65, n.s.). As for the slow minimal pairs and running speech, there is likewise no significant difference either for the short vowels (t (160) = -0.4, p = 0.68, n.s.) or the long vowels (t (147) = -1.2, p = 0.24, n.s.). Turning to the fast pairs and running speech, however, we find that the differences are barely significant for the short vowels (t (160) = -2.0, $p < 0.06$) and quite significant for the long vowels (t (147) = -3.3, $p <$ 0.002).

DISCUSSION AND CONCLUSION

By and large, in answer to the question raised at the beginning of this paper, we can say that the quantity distinction between short and long vowel phonemes in Thai is certainly maintained in a variety of speaking conditions. This is true despite the fact that, except for the slow pairs in Figure 2, the standard deviations for the short and long vowels overlap in the pooled data of Figures 2,3, and 4, giving us to understand that there is a fair amount of overlap between the ranges of values for the two categories in all three speech conditions.[5] On the face of it, there would seem to be a problem in that one could not simply pick a datum at random from the region of overlap in the range of durations, even for a single speaker, and decide with great confidence whether it was from a short or long vowel. We can turn to the best controlled of the conditions, the minimal pairs in carrier sentences, to find at least part of the answer. The analysis of variance shows the choice of word in the set of 16 to be a highly significant factor; it also shows a significant interaction between the factors of word and vowel length. As can be seen in Table 1, there is considerable variability in the phonemic makeup of the eight pairs of words. That is, not only are we dealing with eight different vowels but also quite a variety of consonantal contexts and two tones.[6] Fitting well with this finding is the failure to show the choice of tokens of the words to be significant; thus, utterances of the same word, having, of course, exactly the same phonological composition, do not differ significantly from each other in vowel duration.[7] Of course, the variability in these factors was greater in the excerpts read from a script and even more so in the sample of running speech, although there was no significant difference between the latter two sets of data.

In an attempt to cope with some of these factors, I thought that focusing on single vowel pairs distinguished by length might help. For only one vowel pair, /a a:/, were there sufficient data in the conversation of two speakers for statistical treatment with an unpaired *t*-test. For S.A., with 39 instances of the short vowel /a/ and 33 of the long /a:/, the difference was highly significant: t (70) = -10,5, $p < 0.001$. For TL, with 62 instances of the short vowel and 60 of the long one, the difference was also highly significant: t (120) = -7.3, $p < 0.001$. Thus, there is agreement with the findings for the whole set. At the same time, it must be admitted that looking at a single vowel pair yielded very little reduction of overlap. This is perhaps not surprising when we consider that the items included both CV and CVC syllables; furthermore, all other contextual variables were not under control.

In relaxed running speech other variables must play a role in how carefully separated short and long vowels are. These include tempo, emphasis, familiarity of the subject matter, first or later occurrence of an important word, sentence intonation, position in the sentence, ambient noise, and perhaps other factors, such as variation in vowel quality correlated with length.

Presumably speakers of Thai, as well as of other languages with phonological distinctions in quantity, learn to take these factors into account while processing the relative durations of short and long vowels. That is, as with other kinds of phonemic contrast, the mental grammar may embrace several phonetic correlates of the length distinction, even if one of them, relative duration, is more powerful than the others. The work of Svastikula (1986) certainly supports this contention for the factor of rate.

A question of method arises, namely, whether one could have more control over these variables while still eliciting truly spontaneous speech. Some approaches have been tried that yield better comparability between speakers (e.g., Terken, 1984; Anderson *et al.*, 1991; Swerts & Collier, 1992). Speakers are told, one by one, to do the same verbal task, such as describing a graphic network or reading a map. Such a task makes it highly likely that all the speakers used will produce linguistically similar utterances that are natural responses to the prescribed situation, even though their semantic scope and, probably, their syntactic range are somewhat constrained. My choice of relaxed, lively conversations between people well known to each other on topics of their own choosing bought virtually perfect spontaneity at the price of little or no control over contextual variables. As an extension of this project, it will certainly be desirable to consider eliciting monologues built on carefully constructed situations or tasks.

I plan, following a common practice in experimental phonetics, to seek perceptual validation of this general finding for speech production. The first step will be to present unaltered words from the present samples of running speech to native speakers of Thai for identification. Of course, it will be necessary to choose words that have counterparts of the opposite length. The words chosen will not be cut from the immediately surrounding context, because this could unreasonably mislead the listener. Instead, I will low-pass filter the context to remove syntactic and semantic redundancy, while at the same time keeping the intonational line and tonal features of the context and a speechlike quality. That is, the contexts will be unintelligible while still sounding speechlike. The resulting stimuli will be used in identification tests to determine whether the phonemic contrast is preserved not only in production, as it would seem from the results of the present study, but also in perception. Next, if the distinction *is* perceptually robust, the acoustic structure of the words will be manipulated to see how well the cues in casual speech match those in citation forms. Work by Abramson and Ren (1990)

revealed that spectral differences between short and long vowels in minimal pairs of words embedded in carrier sentences have no more than a small effect on the efficacy of relative duration as a cue. Perhaps the role of this and other features is, under certain conditions, sometimes larger in running speech. The undertaking will be begun with changes in duration of the vowels of the words taken from passages of connected speech. The incrementally lengthened original short vowels and incrementally shortened original long vowels will be used as stimuli in perception tests. The next obvious step would seem to be the introduction of incremental spectral differences by raising and lowering formant frequencies.

REFERENCES

Abramson, A. S. (1962). *The vowels and tones of standard Thai: Acoustical measurements and experiments*. Bloomington: Indiana University Research Center in Anthropology, Folklore and Linguistics, Pub. 20 [= *International Journal of American Linguistics, 28*, No. 2, Part III].

Abramson, A. S. (1974). Experimental phonetics in phonology: Vowel duration in Thai. *Pasaa: Notes and news about language teaching and linguistics in Thailand, 4*, 71–90.

Abramson, A. S. (1987). Word-initial consonant length in Pattani Malay. In *Proceedings of the XIth International Congress of Phonetic Sciences* (Vol. 6, pp. 66–70). Tallinn, Estonia: Academy of Sciences of the Estonian S.S.R.

Abramson, A. S. (1989). Laryngeal control in the plosives of Standard Thai. *Pasaa: Notes and news about language teaching and linguistics in Thailand, 19*, 85–93.

Abramson, A. S., & Ren, N. (1990). Distinctive vowel length: Duration vs. spectrum in Thai. *Journal of Phonetics, 18*, 79–92.

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Sotillo, C., & Thompson, H. S. (1991). The HCRC Map Task Corpus. *Language and Speech, 34*, 351–366.

Barik, H. C. (1977). Cross-linguistic study of temporal characteristics of different types of speech materials. *Language and Speech, 20*, 116–126.

Bennett, D. C. (1968). Spectral form and duration as cues in the recognition of English and German vowels. *Language and Speech, 11*, 65–85.

Dunn, M. H. (1993). *The phonetics and phonology of geminate consonants: A production study*. Unpublished doctoral dissertation, Yale University, New Haven, Connecticut.

Fowler, C. A. (1988). Differential shortening of related content words produced in various communicative contexts. *Language and Speech, 31*, 307–319.

Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language, 26*, 489–504.

Gandour, J. (1984). Vowel duration in Thai. *Crossroads, 2*, 59–64.

Gandour, J., & Dardarananda, R. (1984). Prosodic disturbance in aphasia: Vowel length in Thai. *Brain and Language, 23*, 206–224.

Gandour, J., Weinberg, B., Petty, S. H., & Dardarananda, R. (1987). Vowel length in Thai alaryngeal speech. *Folia Phoniatrica, 39*, 117–121.

Hadding-Koch, K., & Abramson, A. S. (1964). Duration versus spectrum in Swedish vowels: Some perceptual experiments. *Studia Linguistica, 18*, 94–107.

Jones, D. (1950). *The phoneme: Its nature and use*. Cambridge: W. Heffer & Sons.

Kohler, K. J. (1992). Dauerstrukturen in der Lesesprache. *Arbeitsgerichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel, 26*, 225–238.

Laan, G. P. M. (1992). Perceptual differences between spontaneous and read aloud speech. *Proceedings of the Institute of Phonetic Sciences Amsterdam, 16*, 65–79.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, Mass.: MIT Press.

Levin, H., Schaffer, C. A., & Snow, C. (1982). The prosodic and paralinguistic features of reading and telling stories. *Language and Speech, 25*, 43–54.

Liberman, A. M. (1957). Some results of research on speech perception. *Journal of the Acoustical Society of America, 29*, 117–123.

Liberman, A. M., & Cooper, F. S. (1972). In search of the acoustic cues. In A. Valdman (Ed.), *Papers in linguistics and phonetics to the memory of Pierre Delattre* (pp. 329–338). The Hague: Mouton.

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech, 6*, 172–187.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops. *Word, 20*, 384–422.

Remez, R. E., Berns, S. M., Nutter, J. S., Lang, J. M., Davachi, L., & Rubin, P. E. (1991). On the perceptual differentiation of spontaneous and prepared speech. *Journal of the Acoustical Society of America, 89*, 2011–2012. (Abstract)

Saravari, C., & Imai, S. (1983). Perception of tone and short-long judgement of vowel of variants of a Thai monosyllabic sound. *Journal of Phonetics, 11*, 231–241.

Sittachit, K. (1972). ข้อสังเกตเกี่ยวกับสระสั้นสระยาวในภาษาไทย. [Observations on short and long vowels in Thai]. In J. G. Harris, & R. B. Noss (Eds.), *Tai phonetics and phonology* (pp. 28–32). Bangkok: Central Institute of English Language.

Straka, G. (1959). Durée et timbre vocalique. Zeitschrift für *Phonetik und allgemeine Sprachwissenschaft, 12*, 276–300.

Sutadarat, S. G. (1978). A phonological description of Standard Thai. Ann Arbor: University Microfilms (79-19821).

Svastikula, K. (1986). A perceptual and acoustic study of speech rate in distinctive vowel length in Thai. Unpublished doctoral dissertation, University of Connecticut, Storrs.

Swerts, M., & Collier, R. (1992). On the controlled elicitation of spontaneous speech. *Speech Communication, 11*, 463–468.

Terken, J. M. B. (1984). The distribution of pitch accents in instruction as a function of discourse structure. *Language and Speech, 27*, 269–289.

Tingsabadh, K., & Abramson, A. S. (1993). Thai. *Journal of the International Phonetic Association, 23*, 24-28.

FOOTNOTES

[*] In K. Tingsabadh & A. S. Abramson (Eds.), *Essays in Tai linguistics*. Bangkok: Chulalongkorn University Press (in press).

[†] Also University of Connecticut, Storrs.

[1] Formal linguistic criteria may make it convenient to posit gemination, even when no phonetic evidence supports this analysis. An example is the presence of a morpheme or word boundary within the long segment. See Dunn (1993) for data supporting the probability of "unitary" geminates (long consonants) in Finnish but the probability of overlapping articulatory gestures in Italian.

[2] For a brief summary of that early work, see Liberman (1957). A good account of the evolution of the concept is to be found in Liberman and Cooper (1972).

[3] An apparent exception is the set of phonological constraints on syllable types within the language. Since one cannot utter a syllable without invoking such rules, we might argue that we are dealing here for all practical purposes with bottom-up information only.

[4] For seven out of the eight pairs, only words with mid and low tones were used, because they were meant originally for perceptual experiments in which the vowels were to be lengthened or shortened (Abramson & Ren, 1990). These tones are least susceptible to distortion in such an operation.

[5] Inspection of Tables 2 and 3, however, suggests that individuals vary somewhat in the amount of overlap even in such a variety of contexts.

[6] See footnote. 4. Although only the mid, low, and rising tones occur here, all five tones appear quite freely in the excerpts and the running speech.

[7] There were, it is true, only two tokens of each word for each rate for every speaker. Had there been more tokens with the same statistical result, there would be even more support for the internal solidarity of the word.

Haskins Laboratories Status Report on Speech Research
1993, SR-114, 101-112

Calibration, Validation, and Hardware-Software Modifications to the Carstens EMMA System

Peter J. Alfonso,† J. Robert Neely,‡ P. H. H. M. van Lieshout,†††
W. Hulstijn,††† and H .F. M. Peters†††

INTRODUCTION

This paper is a report of the experiences gained over a period of about two and a half years by a working group from the University of Nijmegen and the University of Illinois with the Carstens Articulograph AG100 articulatory magnetometer system, herein referred to as the Carstens EMMA (Electro-Magnetic Midsagittal Articulometer) system. The initial experiments were conducted at the University of Nijmegen using a 1990 version of the Carstens EMMA system. A second Carstens EMMA system was brought on-line at the University of Illinois in 1992. One of the original aims of the collaborative research was to examine the stability of measures of spatial and temporal coordination of the supralaryngeal speech structures over time and across speech rate for a large number of control speakers and speech-disordered populations. It became clear early on that the Carstens commercially available software and certain components of the hardware would not suffice for the purposes of the proposed experiments. Early experience indicated also that calibration and validation protocols had to be developed to make meaningful comparisons: 1) across a large number of control and experimental subjects, 2) within subjects across time sessions, and 3) across the two Carstens EMMA systems. In Sections II through IV below, the hardware and software modifications and some of the protocols that were developed to carry out experiments of the type described here are presented. In the first section an example of the nature and demands of the experiments is given; in the last section a small portion of the results is described briefly.

I. Description and Requirements of the Experiment

The objective of this example experiment was to determine the stability of measures of spatial and temporal coordination of the supralaryngeal speech structures over time and across normal, fast, and slow speech rates for a large number of control speakers and adult stutterers. The measures of interest included motor equivalence covariability, relative-timing, and the temporal order (sequence pattern) of the movements of the upper and lower lips, tongue blade, and jaw. A reference receiver coil was attached to the nose for error correction in the event of head movement within the helmet. Twenty subjects, 10 per group, were run. For each session, subjects were asked to produce 20 perceptually fluent repetitions of the target words /pap/, /sas/, and /tat/ embedded in the Dutch carrier phrase "Zij zei CVC alveer." The 60 phrases were blocked by rate and produced first at a normal speech rate, than again at a fast speech rate, and finally at a slow speech rate. Sessions were repeated three time, and the interval between sessions was about two weeks. Total stimuli collected during the experiment included some 10,800 phrases (20 subjects × 3 target words × 20 repetitions × 3 rates × 3 sessions). The analog speech acoustic signal was recorded and was later digitized and time-aligned with the EMMA signals.

During the course of the pilot studies it became clear that modifications to the original helmet design would have to be made for at least two reasons. First, a single session took about 45 to 60 minutes and it became evident from the comments of the subjects that the weight and general comfort of the original helmet was problematic for runs of this duration. In addition, the stability of the head within the helmet was also a concern for

---

Research supported in part by Fulbright Research Award to the first author, NIH Grant DC-00121 to Haskins Laboratories, and University of Illinois Research Board Grant.

runs of this duration. Second, the precise repositioning of the helmet and receiver coils across sessions was essential to the success of the project and the original helmet design did not provide a mechanism for which minor vertical and horizontal adjustments could be made nor did it provide a mechanism for which reference to anatomical landmarks could be made easily.

Early experience with the Carstens EMMA system indicated that equipment calibration procedures would have to be developed. Hardware and software would also have to be developed to rotate the data to the occlusal plane or other points of reference. Finally, the pilot studies indicated the need for development of subject and data validity criteria.

The large amounts of data collected in this project exceeded the potential of the manufacturer's software and PC's of the time. Thus, significant software development was required for the purposes of data transport to more powerful computer systems, automatic data pre-analysis routines (e.g., smoothing, rotation, etc.), and efficient data analysis routines (e.g., trajectory display of the x and y components of the displacements and velocities of the movements over time).

Thus, the requirements of the project were such that modifications of some of the hardware supplied by the manufacturer at the time and development of new hardware and software would have to be completed. Furthermore, calibration and data validation protocols and certain operational criteria were needed.

II. Hardware Modifications

II.A. Helmet modifications. At the time of the original experiments, only one helmet design was available from the manufacturer. At the time of the Munich conference, three helmets were available from the manufacturer. One of the three, referred to as the Nijmegen helmet, was developed as a result of the experience gained with the project described here.

As alluded to above, the original helmet design was a concern in regard to the requirements of the project because of its weight and the considerable duration of a single session, and because it lacked a way in which minor position adjustments could be made easily, particularly in regard to the consistent positioning of a single subject across sessions. Figures 1 through 5 show the modifications that were made to the original helmet design to address these concerns. Figure 1 shows the complete Nijmegen helmet in place. Note that an adjustable counter-weight system is provided and that it mounts to the rear of the subject's chair. The counter-weight system is attached to the helmet by the lower pair of two horizontal bars manufactured directly to the helmet. The right side of the Nijmegen helmet is depicted in Figure 2 and shows an upper and lower horizontal bar. The upper pair of horizontal bars attaches directly to the adjustable head mount shown in Figures 3 and 4. Adjustment mechanisms are provided so that the head mount attaches securely to the subject's head.

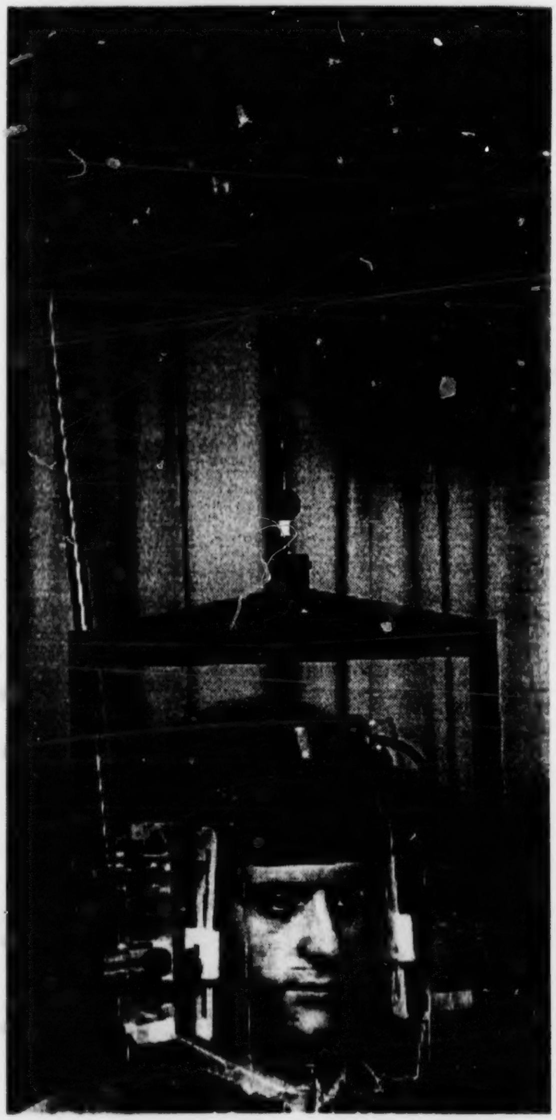


Figure 1. Carstens helmet with Nijmegen extension showing the modified helmet and counterweight system in place.

BEST COPY AVAILABLE

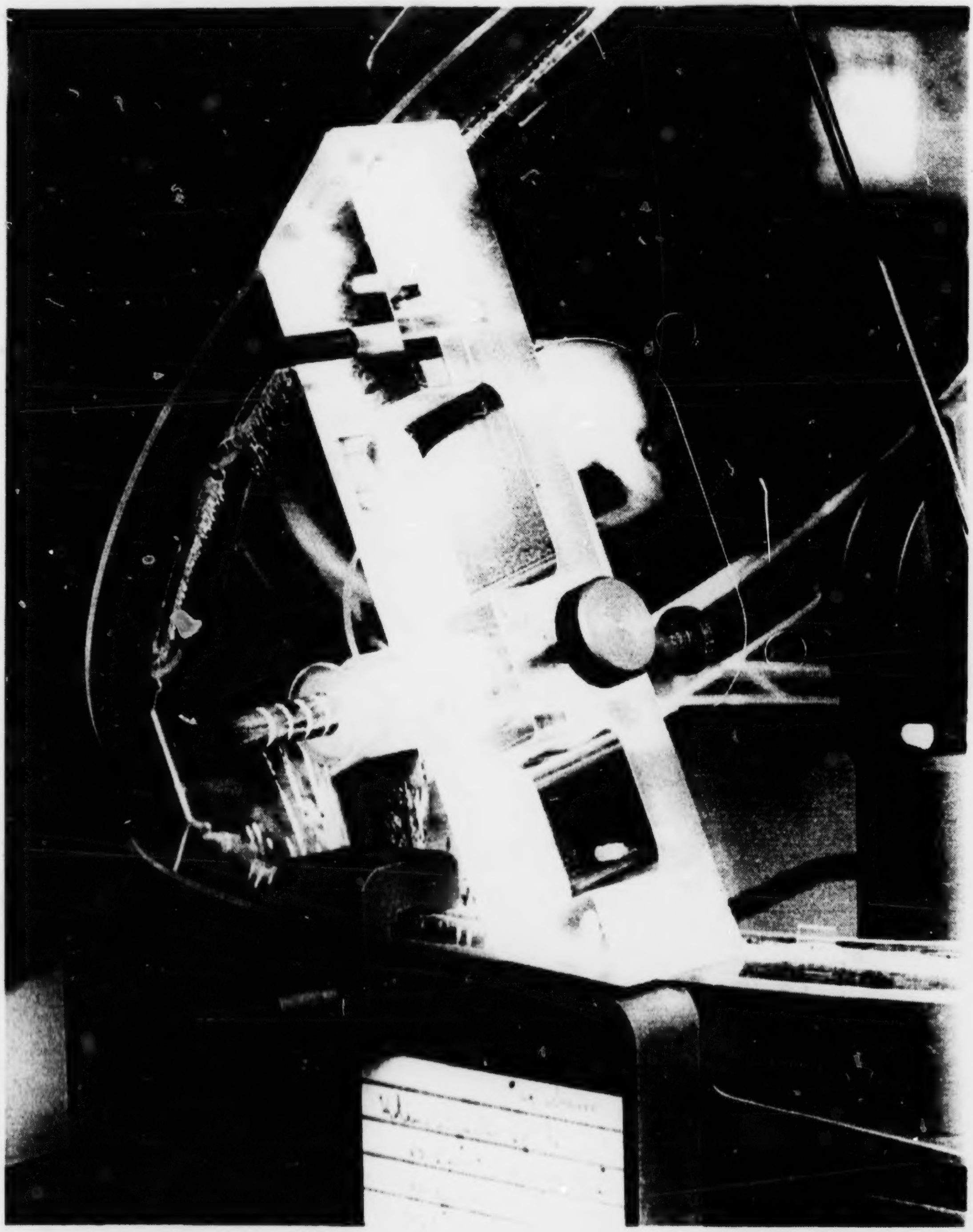


Figure 2. **The right side of the Nijmegen helmet showing upper and lower horizontal attachment bars. The upper bar attaches directly to the head mount shown in Figure 3. The lower bar attaches directly to the counterweight system.**

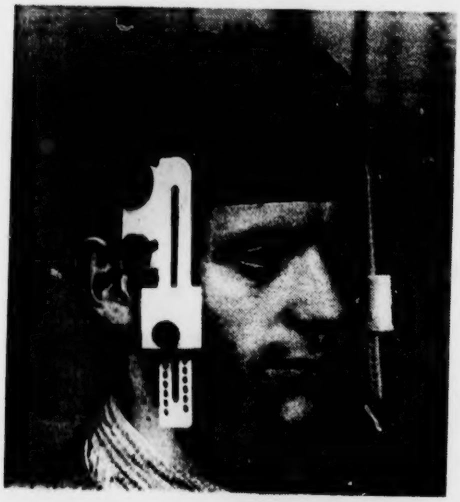


Figure 3. **Adjustable head mount which is provided as part of the Nijmegen helmet extension.**

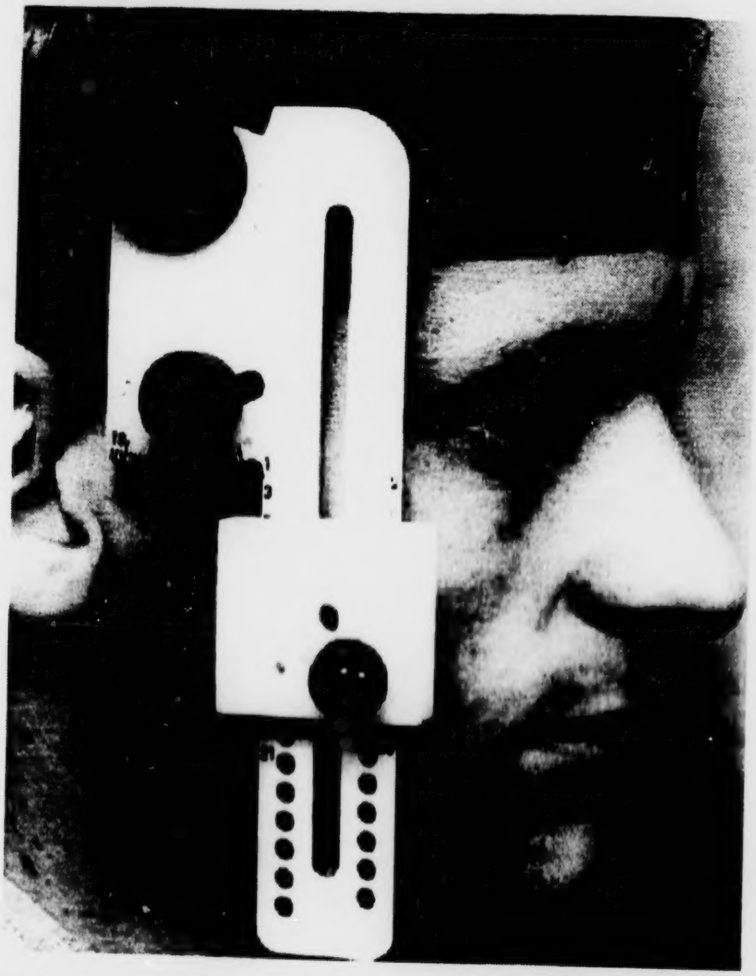


Figure 4. **Side view of the head mount showing mechanism for rotational and vertical adjustments.**

The attachment points between the head mount and helmet are such that the position of the subject's head relative to the helmet can be rotated or displaced vertically. A detailed view of the in-place counter-weight system, helmet, and head mount is shown in Figure 5. Further details of the Nijmegen helmet design can be obtained from the Carstens company or from the University of Nijmegen.

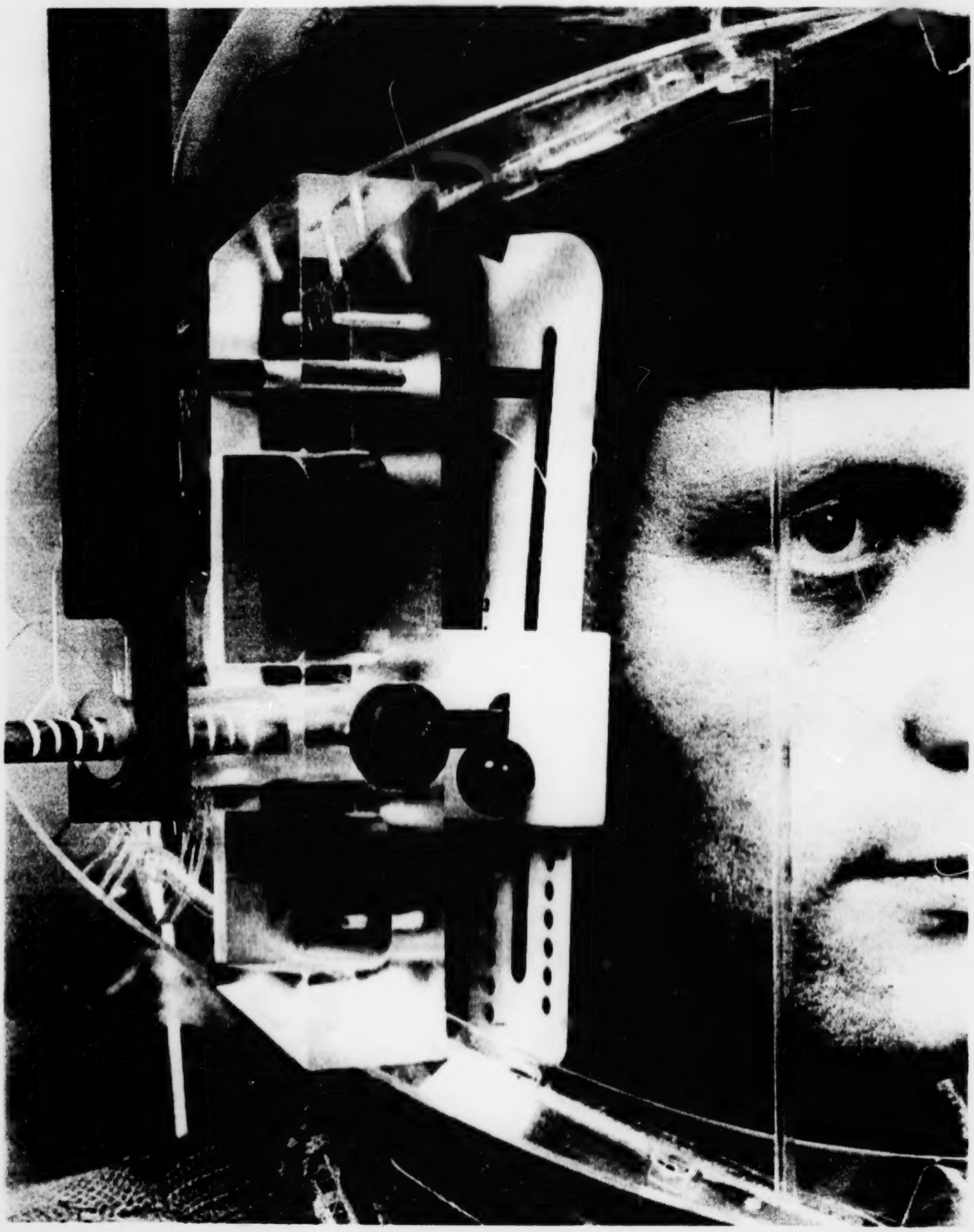


Figure 5. **Detailed view of the in-place Nijmegen helmet, head mount, and counterweight bracket.**

BEST COPY AVAILABLE

II.B. Software and hardware development for rotational reference and spatial calibration. In order to make meaningful comparisons among Carstens EMMA data and other physiological data on the supralaryngeal vocal tract, it was necessary to rotate the Carstens EMMA data to common planes of reference. Although there are a number of ways that rotation can be accomplished, the early pilot work showed that acceptable results were obtained by reference to the nose reference coil and the jaw coil. Using a standard rotation algorithm, the EMMA data were rotated around the center point of the helmet so that the nose to jaw angle was 0 degrees on the Y axis. Since a new rotation angle was computed for every utterance, the angle can be used as a check against erroneous head movement within the helmet. Differences among rotation angles during a single run of about 1 degree or less were observed typically, indicating that head slippage within the helmet was not a major concern. To aid in the precise repositioning of the helmet across sessions, records of the surface points of the receiver coils were kept and comparisons of the rotation angle were made across sessions. Rotation angles across the different subjects varied between 10 to 30 degrees as a function of best helmet fit and individual head characteristics.

Another method, one borrowed from x-ray imaging, is to record the position of a structure that lies parallel to the occlusal plane. Figure 6 shows an example of a device in use for this purpose at Haskins Laboratories with the MIT EMMA system developed by Perkell and his colleagues (Perkell, Cohen, Svirsky, Matthies, Garabieta and Jackson, 1992). Prior to the start of data collection, two receiver coils are placed on the midline of the device. The subject is instructed to bite on the plate and the position of the two alignment coils are recorded. EMMA data collected during the run are rotated to parallel the alignment coils on the X axis. The known distance between the two alignment coils can be used as a spatial calibration reference with software analysis routines.

At the present time, the manufacturer provides software for data rotation. However, the experimenter must decide on the method of reference. The advantages of standardization could best be served by users of the Carstens EMMA system by uniform implementation of a device similar to that depicted in Figure 6.

III. Calibration and Validity Criteria

This section discusses some of the practical issues that arose during the course of Nijmegen-Illinois project with the two Carstens EMMA systems. The intent is to point out some of the problems that were encountered and to share the solutions that were devised, as well as to point out a few problems that continue to exist. Many of these same issues were discussed in detail by other conference presenters, particularly Drs. Honda, Hoole, Nye, Perkell, and Schonle, and since their reports appear elsewhere in these Proceedings, the issues need not be discussed in great detail here.

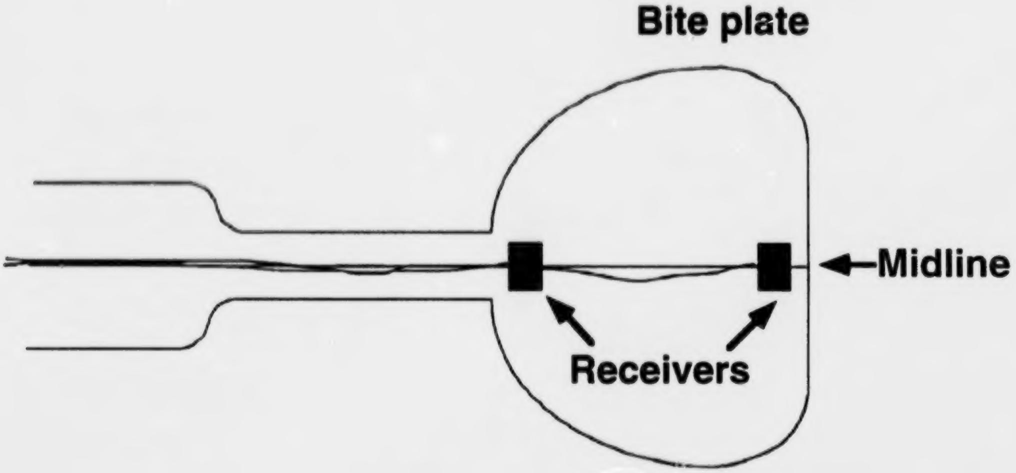


Figure 6. **Bite plate with two receiver coils. The position of the bite plate coils is used to reference the occlusal plane for subsequent rotation of data captured during the run.**

III.A. Equipment Calibration

III. A.1. Transmitter (DC offset) stability. The manufacturer recommends that the equipment be turned on for a period sufficient for the three transmitters coils to reach equilibrium before measurements are attempted. However, it was observed that transmitter stability, by observation of the DC offsets, may not have been reached even after warm-up periods of two hours or greater. Since the heat generated by the transmitters and their amplifiers should have stabilized well before two hours and since the magnitude of the shifts in the transmitter output voltages were comparable across the Nijmegen and Illinois EMMA systems, a question arose as to the inherent internal stability of the system. However, repeated measures across five minute intervals of the static position of the five receiver coils at warm-up periods of one hour or longer indicate that the effect of the apparent transmitter instability on immobile receiver coil position identification is negligible. Average measurement error in the range of .1mm to .2mm. was detected with both EMMA systems. Apparently, the tilt correction software compensates for drifts in the transmitter output voltages. Thus, the results of the informal tests are that the system reaches sufficient levels of stability to allow collection of positional data as a function of time, although the time course of the apparent positional stability is not clear at the moment and requires further examination. Finally, system stability should be ascertained before transmitter and receiver calibration procedures are attempted since drifts in transmitter output voltages during calibration procedures could result in serious measurement error. Dr. Nye, in a paper appearing elsewhere in these Proceedings, discusses the thermal stability of the MIT EMMA system in operation at Haskins Laboratories.

III. A.2. Static calibration measurements. Until the effects of transmitter stability are known better, particularly stability as a function of time, a simple procedure to verify the spatial position of the five receiver coils was followed immediately before and after data recording sessions. Immediately after the transmitter and receiver coils were calibrated following the manufacturer's instructions, the receiver coils were carefully placed in the holder supplied by the manufacturer and their known physical position was recorded and measured using the XHADES software discussed below. The average measured distance across the receiver coils was used as the calibration reference with the XHADES software program. Thus, the procedure can be used to verify that the system is maintaining equilibrium and as a source of calibration reference data for analysis software routines.

Dr. Honda has completed intensive bench tests of the static positional resolution of the Carstens EMMA system and some of this work is reported elsewhere in these Proceedings. The results of the Nijmegen-Illinois calibration tests compare favorably with the results of the extensive work completed by Dr. Honda.

Dr. Hoole reports elsewhere in these Proceedings on an improved procedure to calibrate the transmitter and receiver coils of the Carstens EMMA system. Immobile coil stability, following the Hoole calibration procedures, is improved to about .03mm.

III. A.3. Dynamic calibration measurements. Ideally, the Carstens EMMA system should be calibrated with a dynamic device since the goal of the system is to capture articulatory motion. At the time of the Munich conference, a custom circle calibration device similar to that described by Perkell et al. (1992) or other devices similar to those described by Honda or Nye appearing elsewhere in these Proceedings were not available by the manufacturer. A custom made device is essential so that appropriate calibration data of the type reported by Perkell and Nye can be collected easily before every data collection session. At the present time, the Carstens group is manufacturing a dynamic calibration device and information on the calibration hardware and software is forthcoming.

III.B. Subject Calibration

III. B.1. Subject acceptance criteria. In the course of running a large number of control and experimental subjects, it became clear that not every subject was a good candidate for the EMMA protocols reported here. For example, subjects who do not exhibit significant asymmetrical lateral or vertical tongue movements are preferred (see, for example, Stone, Faber, Raphael, and Shawker, 1992). Tongue asymmetries were less problematic in regard to tracking errors across most subjects when the phonetic context of the target syllables was restricted to low vowels, e.g., /a/, rather than high grooved vowels (e.g. /i/) and with receiver coil placement locations at anterior rather than posterior tongue sites. With respect to the experimental subjects, a few severe stutterers exhibited lip, tongue, and jaw repetitions at very high rates. Measurement of articulator movements during high rate repetitions for these subjects were not reliable. However, see the paper by Dr. Schonle appearing elsewhere in these Proceedings who reports success with a variety of speech disordered populations.

Most of the research to date on the hazards of electro-magnetic fields (EMF) have implicated frequencies lower than those generated by EMMA systems. However, more recent EMF studies on MRI and CRT instruments, and some common home appliances, suggest that EMMA output frequencies and local field strengths present minimal risk to humans, particularly at the relatively short exposure periods associated with EMMA data collection. However, it would be prudent to monitor this line of research. Until more is known about EMF generated by EMMA systems, the Nijmegen-Illinois project is restricted to adult subjects. See Perkell et al. (1992) for further discussion of this topic.

III. B.2. Static calibration measurements. After the receiver coils were attached to the subject and immediately before and after data were collected, the static receiver coil positions were recorded and measured using the manufacturer's software, and later when it became available, using the XHADES software. The reference (nose) coil to jaw coil angle was used for data rotation (see II.B. above). The computed spatial positions of the five coils can be compared to the surface points of the receiver coils as an informal check of system stability.

A palate trace was recorded immediately before and after data were collected. The method of choice was to ask the subject to trace the palate with the anterior tongue receiver coil. Alternatively, the experimenter can trace the subject's palate by attaching a receiver coil to the experimenter's finger but this method may induce interference with the transmitter outputs. (Care should be exercised to remove watches, rings, and the like when in the immediate vicinity of the transmitter coils. See the paper presented by Nye appearing elsewhere in these Proceedings on the subject of environmental interference.) Comparison between the pre- and post-session palate trace gives an indication of head slippage within the helmet and consequently the suitability of the data within the session.

III. B.3. Dynamic calibration measurements. Immediately prior to the collection of the phrase length stimuli, the subjects were instructed to produce multiple repetitions of the CV syllables /pa/, /sa/, /ta/ and of the CVCVCV syllable /pasata/. Recall that the target CVC words imbedded in the phrases were /pap/, /sas/, and /tat/. The articulator movements were observed using the real-time display provided by the manufacturer's software. The movements were compared informally across repetitions as a final check of the proper operation of the EMMA system. The calibration articulator movements and the corresponding acoustic signals were recorded and analyzed later as part of the data validity criteria described below.

III.C. Data Validity Criteria

The limitations of the new EMMA technology require that inappropriate data, which result from occasional errors induced by factors such as but not limited to misalignment of the transmitter and receiver coils, excessive receiver coil tilt associated with tongue grooving or displacement asymmetry, loose attachment of the receiver coils to the flesh points, be identified and eliminated from the data corpus.

III. C.1. Statistical criteria. The preferred data reduction method was to identify significant outliers in the data set. For example, the central tendency and variability for the vertical displacement for the anterior tongue receiver coil associated with /t/ closure were estimated from the isolated /ta/ repetitions referred to in Section III.B.3. and the /tat/ target syllables embedded in the phrase length stimuli. Similarly, statistical criteria based on tilt correction angles can be employed. Since the 1990 version of the Carstens EMMA system in use at the time of the data collection did not provide access to the correction angle, it was not possible to compare the results of the kinematic procedure with the correction angle procedure. The current version of the manufacturer's software makes the tilt correction angles available to the experimenter.

Other methods that can be used to identify spurious data include the reference rotation angle referred to in Section II.B., analysis of the reference coil position as a function of time,[1] comparisons among the pre- and post-session attached and immobile receiver coil positions and the palatal traces referred to in Section III.B.2., and the obvious discontinuities in the movement profiles per utterance that can be observed by the real-time display provided by the manufacturer.

III. C.2. Need for standardized data validity criteria. Procedures other than those presented here will no doubt be proposed as use of EMMA technology becomes widespread throughout the field. It should be clear, however, that a number of control factors must be addressed in order to assure the accuracy and validity of articulatory movement data captured by EMMA systems. The field would be well served if standardization of procedures and criteria to address these control factors, some of which are presented here in the Section III.C., could be reached.

IV. Software Modifications

The large amounts of data collected in projects of this type exceed the potential of the manufacturer's analysis software and the storage capability of the PC's running the EMMA hardware at the Universities of Nijmegen and Illinois. Thus, significant software was developed to transport the data to a VAX computer, to automatically perform "pre-analysis" routines such as rotation and smoothing, and to efficiently analyze large quantifies of kinematic data.

At the time that the original data were collected in 1990-91, the Carstens EMMA system did not provide the capability to digitize and align the speech signal with the EMMA signals. Thus, the analog speech signal was recorded and then digitized after the session. Software was developed to time align the digital speech signal with the EMMA signals. The current version of the Carstens EMMA system supports a two-channel A/D speech input; some of the pre-processing and analysis routines mentioned below are also provided currently by the manufacturer.

IV.A. Data Transport Routine

The original MS-DOS EMMA files were transferred to a VAX station 4000 machine via a standard FTP. Each file was then converted to VMS format using a public domain program entitled FILE.EXE. Since XHADES requires PCM formatted data, it was necessary to demultiplex each VMS file into 10 separate PCM files (five receiver coils × two dimensions). Four additional PCM files were created later and are described below. Although demultiplexing is completed at this stage, the PCM headers require information about the form of the data. Some of the data information is derived from the pre-analysis routines; thus, the EMMA PCM formatted files are not created until pre-analysis is complete. The structure of the speech acoustic files are far less complicated and the VMS headers of the speech files are converted to PCM format at this stage.

IV.B. Pre-analysis routines

IV.B.1. Data rotation. The data points were rotated so that the reference nose receiver coil and the jaw receiver coil correspond to a zero angle along the Y axis. A rotation angle is computed for each phrase length stimulus using a standard rotation algorithm where:

$$x'' = x + (x - x')\cos(\text{theta}) - (y - y')\sin(\text{theta})$$
$$y'' = y + (x - x')\sin(\text{theta}) + (y - y')\cos(\text{theta})$$

The data were rotated around the center point of the helmet. Rotation angles of 10 to 30 degrees were derived for the twenty subjects who took part in the project.

IV.B.2. Filtering. The data shown in Section V below were loss-pass filtered at approximately 30 Hz using an eleven sample triangular window. The program allows for selection of various window sizes.

IV.B.3. Derived independent tongue and lower lip signals. The jaw receiver coil signal is subtracted from the tongue and lower lip receiver coil signals in each displacement dimension, in effect creating four new files. Thus, a single PC formatted EMMA file is converted to fourteen PCM formatted files. The derived independent tongue and lower lip signals are useful, for example, in calculating the independent contribution of the jaw and tongue toward a lingual constriction. Examples of this type of analysis are shown in Figure 7 and Table 1.

IV.B.4. Derived velocity. The first derivative is computed for each of the fourteen files using a standard differentiation algorithm.

IV.B.5. Data normalization. PCM format requires that the data be normalized into 12 bit integer values. To this end, the normalization routine identifies the maximum value in Carstens units for each of the 14 files and using a standard linear regression translation algorithm (where $y = mx + b$) converts the original data into PCM normalized values. XHADES can translate the normalized values back to absolute space coordinates since the algorithm values of m and b are included in the PCM header.

IV.B.6. Output to PCM files. The pre-analyzed data are output as 14 separate PCM formatted files.

IV.C. XHADES Analysis Routines

Among the advantages of the XHADES Haskins Analysis program is the executive code language SPIEL that allows for the automatic initiation of a number of sequential procedures. Figure 7 shows a screen dump representing the production of one of the phrases summarized in Table 1. XHADES interactive routines and SPIEL commands were used to accomplish the following:

1) Display the eight appropriate files. The three upper-most records on the left hand side of the figure represent the vertical displacement of the anterior tongue receiver coil, which corresponds to the net tongue-jaw displacement, the derived independent tongue signal, and the jaw receiver coil as a function of time. The bottom-most record is the speech acoustic signal for the utterance "zij zie

tat alveer." The three upper-most records on the right side of the figure represent the corresponding velocity profiles.

2) Segment the acoustic signal. Figure 7 shows that phrase duration was measured as the period from the onset of voicing in the phrase-initial vowel to the onset of voice in "alveer."

3) SPEIL commands set labels at zeros and peaks of the velocity functions corresponding to /t/ closure in /tat/, transfer labels to the displacement files, calculate the closure displacements for each of the three records, and output to the appropriate data files. See Rubin, MacEachron, Tiede, and Maverick (1991) for a detailed description of the XHADES program.

V. Results of the Experiment

A portion of the results is shown in Table 1 for the example experiment described in Section I. The upper portion of the Table shows the average vertical displacements in mm of the jaw receiver coil (J), derived independent tongue signal (T), and the tongue receiver coil (J+T) for initial /t/ closure in the target syllable /tat/ for about 20 repetitions of the carrier phrase "zij zie tat alveer." Data are shown for four subjects and across two sessions. The lower portion of the Table shows the corresponding average phrase durations as measured in Figure 7. A major point to be made here is that none of the EMMA data represented in Table 1 were subjected to the data validity criteria discussed in Section III.C.; however, the data appear reasonable even before this condition was satisfied. First, the standard deviations are relatively small compared to the means. Second, the jaw-tongue synergies shown in session one are also shown in session two. For example, subject one shows more jaw than tongue displacement for lingual constriction in both sessions. Alternatively, subject two shows more tongue than jaw displacement in both sessions. Third, in the majority of cases the magnitudes of the displacements are reasonably similar across sessions. A notable exception is the combined J+T displacement for subject four. Thus, Table 1 represents the worse case solution in that none of the data were discarded on the bases of the criteria discussed in Section III.C., yet for the most part the data appear appropriate. Four of the approximately 80 utterances represented in Table 1 were later excluded as a function of the data validity criteria discussed in Section III.C.

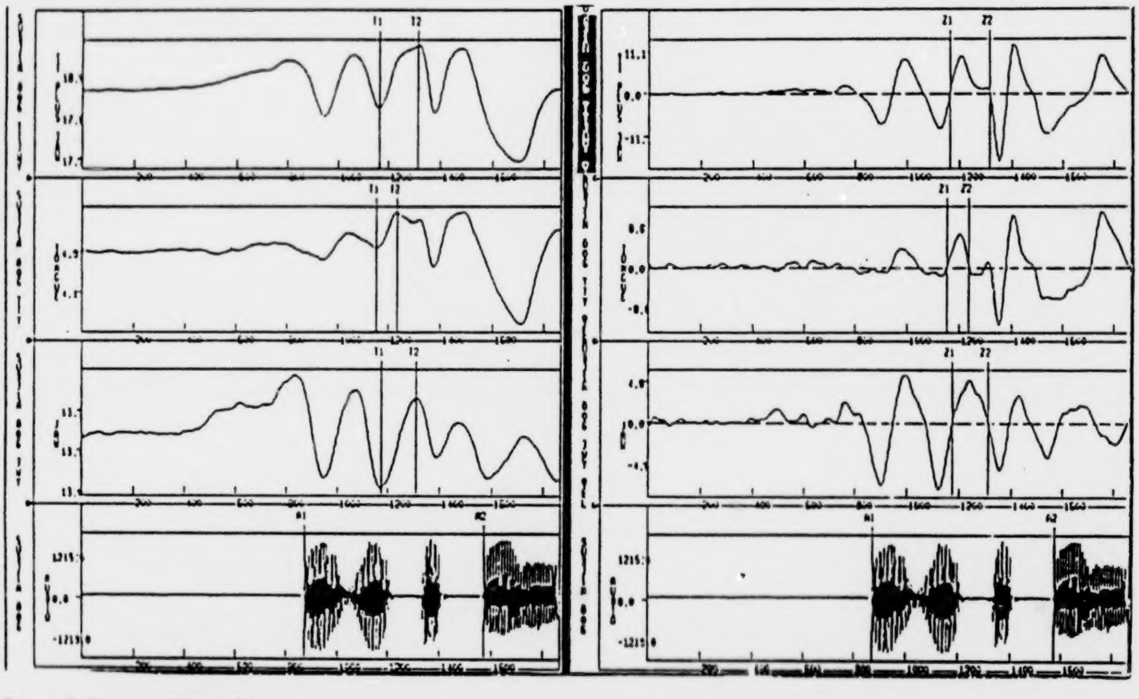


Figure 7. Typical XHADES display showing the Dutch utterance "Zij zie tat alveer." Upper three records on the left represent the vertical component of the tongue plus jaw, derived tongue, and the jaw displacements. The upper three records on the right represent the corresponding velocity signals. The lower records represent the acoustic signal.

CONCLUSIONS

Although the processes involved in the monitoring of articulatory movements with the Carstens EMMA system are not straightforward and certain factors must be controlled to insure valid results, the data reported here and in the reports by Drs. Honda and Hoole appearing elsewhere in these Proceedings indicate that the resolution of the Carstens EMMA system to monitor the structures of speech located within the head is at least equal to that obtained by x-ray microbeam tracking.

The original experiments of the Nijmegen-Illinois project indicated that the 1990 version of the Carstens EMMA system lacked certain hardware and software systems that were necessary to insure that some of the critical factors that are necessary to obtain valid data of the types described here could be controlled. Thus, it was necessary to develop certain hardware and software systems to supplement what was commercially available at the time. However, the manufacturer has made available a number of improvements in the past three years, more improvements are in the development stage, and he has cooperated with users to meet their individual needs.

REFERENCES

Perkell, J. S., Cohen, M. H., Svirsky, M. A., Matthies, M. L., Garabieta, I., Jackson, M. T. T. (1992). Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. *Journal of the Acoustical Society of America, 92(6)*, 3078-3096.

Rubin, P., MacEachron, M., Tiede, M., Maverick, V. (1991). Haskins Analysis, Display, and Experiment System (HADES). Haskins Laboratories Internal Memorandum, October, 1991.

Stone, M., Faber, A., Raphael, L. J., Shawker, T. H. (1992). Cross-sectional tongue shape and linguopalatal contact patterns in [s], [S], and [l]. J. *Phonetics, 20*, 253-270.

FOOTNOTES

[†] Also The University of Illinois at Urbana-Champaign.

[‡] The University of Illinois at Urbana-Champaign.

[†††] The University of Nijmegen, The Netherlands.

[1] At the time of data collection, the available Carstens EMMA system permitted the monitoring of only five receiver coils. Thus, the number of available reference coils was limited. Currently, the manufacturer provides a 10 receiver coil option permitting the allocation of two or more reference coils. Other methods, such as Selspot instrumentation, are possible to monitor head movements within the helmet.

Haskins Laboratories Status Report on Speech Research
1993, SR-114, 113-138

Morphological Analysis and the Acquisition of Morphology and Syntax in Specifically-Language-Impaired Children

Karen M. Smith-Lock†

In order to find out whether specifically-language impaired (SLI) children show a deficit in the acquisition of inflectional morphology but not syntax, SLI children (mean age 6;2) were compared with language-matched (mean age 4;0) and age-matched controls on their production of passives. Passives were elicited from all groups, with no syntactic errors. Morphological errors were frequent and involved overgeneralization. Morphological skills were further investigated with a series of morphological analysis tasks. The SLI children performed significantly worse than their age-matched peers and were indistinguishable from their language-matched peers. It is concluded that SLI children show proficiency in syntax and deficits in morphology and that morphological analysis skills develop hand in hand with oral language.

The language of specifically-language impaired (SLI) children has been the issue of much recent debate. The debate has focussed on which components of language structure and/or processes are impaired, and in what manner (Clahsen, 1989; Gopnik, & Crago, 1991; Guilfoyle, Allen, & Moss, 1991; Leonard, 1989; Leonard, Bortolini, Caselli, McGregor, & Sabbadini, 1992; Leonard, Sabbadini, Volterra, & Leonard, 1988; Rice & Oetting, 1991). These questions are of interest, not only with respect to clinical issues of identification and remediation of SLI, but also with respect to furthering our understanding of language acquisition in general. The goals of this paper are to examine the relative strengths and weaknesses of SLI children in the domains of syntax and morphology, to explore a possible account of their deficits and to consider the implications for normal language acquisition.

"Specifically language-impaired" (SLI) children, have linguistic deficits in spite of normal non-verbal intelligence, adequate environmental stimulation, normal hearing and lack of identifiable neurological deficits. Specific language impairment is generally diagnosed by comparing a child's level of oral language development to linguistic norms for children her age, as well as to the child's own development in other areas. If a child's linguistic development is not what would be expected for her age, (i.e., if the child's performance falls more than one standard deviation below the mean on standardized tests (McCauley & Swisher, 1984)) and if other areas of development are proceeding normally, a diagnosis of SLI is given.

SLI children typically begin to talk later than normal children and have a low mean length of utterance (MLU)[1] for their age. SLI children acquire grammatical morphemes in the same order as normal children (Johnston & Schery, 1976). However, they typically omit grammatical morphemes at a higher level of language

---

This work was supported by the National Institute of Child Health and Human Development grant HD01994 to Haskins Laboratories, The University of Connecticut Research Foundation and a Natural Science and Engineering Research Council of Canada postgraduate fellowship.

I would like to thank all the students, teachers and speech-language pathologists who participated in the study, as well as Nicola Warrick for her help in generating referrals and Alexandra Gottardo for helping with data collection. Special thanks go to Ignatius Mattingly, Hyla Rubin and Diane Lillo-Martin for their extensive and invaluable assistance.

development (measured in MLU) than do normal children (Johnston & Schery, 1976; Steckol & Leonard, 1979). In spite of this, SLI children do appear to use such morphemes at the early language levels (Johnston & Schery, 1976). Thus, for SLI children, there appears to be a greater delay in the time from first appearance of a morpheme to consistent use of the morpheme.

The fact that SLI children begin to use inflectional morphemes consistently at a higher MLU than normal children suggests that some components of their grammar develop at a more normal rate than others. MLU is not a detailed enough measure to indicate which components are developing ahead of others. Nevertheless, if inflectional morphology is not being used consistently, it may be that the development of more lengthy and complex syntactic structures is responsible for the increase in MLU. This might indicate that SLI children have difficulty with inflectional morphology, but not syntax.

There is some preliminary evidence that syntax is a relative strength for SLI children (Clahsen, 1989; Smith, 1992). Clahsen (1989) proposed that German-speaking SLI children's syntax was intact and that apparent difficulties with syntax could be attributed to morphological deficits. In English, Smith (1992) elicited complex wh-questions (e.g., *What do you think is under the box? Who do you think ate the french fries? How do you think the lady caught the bug?)* from normal children (aged 2;10 to 4;6) and SLI children (aged 3;1 to 5;10). She found that SLI children were able to produce long distance wh-questions at the same age as normal children, as young as 3 years 1 month. Unlike the normal children, some of the SLI children who produced these questions had not yet fully mastered verbal inflections, auxiliary and copula verbs, and do-support, suggesting that their syntactic knowledge was more advanced than their morphological knowledge.

There appear to be (at least) two different phenomena to account for in SLI: the overall delay in language development (and thus, a delay in the first use of inflectional morphemes) and the protracted period of time between first use and consistent use of a particular inflection. The first implies a delay in the acquisition of grammatical competence, the second, a further delay in grammatical performance.

Possible Explanations of SLI

In attempting to explain SLI, several researchers have suggested that SLI children suffer a deficit in their innate linguistic knowledge (Clahsen, 1989; Gopnik, 1990a; Gopnik, 1990b; Gopnik & Crago, 1991; Guilfoyle, Allen, & Moss, 1991; Rice & Oetting, 1991).

Gopnik (1990a, 1990b) and Gopnik and Crago (1991) argued that the grammars of SLI individuals lack features such as aspect, number, gender and the mass/count distinction. In normal speakers, these features, with their phonological representations, are stored separately in the lexicon and added to words when appropriate. Gopnik argues that SLI individuals have no such features, and thus, must store both *cat* and *cats,* with no labelling of the *-s* as a plural marker. She argues that they learn morphologically complex items as unanalysed wholes on an item-by-item basis.

This view predicts that SLI children should not overgeneralize regular endings to irregular forms as normal children do (e.g., *mans* for *men*, *drived* for *drove*) since such a generalisation requires the knowledge of a number or tense feature and the productive application of a rule to new words. Furthermore, SLI children should not be able to comprehend inflections on nonsense words, since they would be unable to recognize the inflectional morpheme representing the feature plural and use a general morphological rule to comprehend the word. Gopnik's proposal implies that SLI children and adults have a deviant grammar due to a deficit in their innate linguistic endowment; they lack morphological features.

An alternative view, proposed by Leonard (1989) and Leonard, Sabbadini, Volterra & Leonard (1988), is that a deficit in the SLI children's perception of the speech signal causes the linguistic input to be filtered or distorted. They found that Italian SLI children showed better ability with several inflectional morphemes than comparable English SLI children and claim that this difference is due to the fact that, in Italian but not English, the inflections are stressed, syllabic, and end in a vowel. Thus, they propose that SLI children have difficulty in perceiving "low phonetic-substance morphemes" (the "surface account"). Low phonetic substance morphemes are "nonsyllabic consonant segments and unstressed syllables, characterized by shorter duration than adjacent morphemes, and, often, lower fundamental frequency and amplitude," such as the tense markers /s/ and /d/ in English. Leonard (1989) and Leonard et al. (1988) propose that this perceptual deficit, combined with the difficulty of building grammatical paradigms (such as those necessary for tense marking), results in the delayed acquisition of grammatical morphemes in English SLI children. The "surface account" offers an account of cross-linguistic data as well as an explanation of SLI children's difficulty with a variety of unstressed grammatical markers.

A perceptual deficit must necessarily affect the perception of non-morphophonemic low-phonetic substance elements as well. Leonard proposes that this can account for production difficulties such as final consonant deletion and weak syllable deletion which appear to occur more frequently in the speech of SLI children than in normal children matched for articulation ability (Ingram, 1981). A perceptual deficit account, however, must be able to explain how SLI children are nevertheless capable of speech perception in general, since much of the speech signal is unstressed and nonsyllabic.

The surface account predicts that SLI children will have difficulty with the acquisition of passive structures. Pinker (1984) proposes that children acquire these structures by using grammatical markers (i.e., *by*) as structural cues. If such grammatical markers are low-phonetic substance morphemes, Leonard points out, the acquisition of the passive will be problematic for SLI children as they will be unable to correctly parse the structure. Although there is some evidence that passives are difficult for SLI children (Menyuk & Looney, 1972), such a finding is not consistent with the observations made above that syntax is a relative strength for SLI children.

Linguistic Analysis Hypothesis

The purpose of this paper is to explore another possible account of the acquisition profile of SLI children, specifically, the Linguistic Analysis Hypothesis. This suggests that SLI children receive adequate linguistic input and have an intact grammatical mechanism but have difficulty analysing the input so that it is available to the grammatical mechanisms. According to this view, the difficulty with inflectional morphology could be due to difficulty analysing morphological structure.

A deficit in linguistic analysis, specifically morphological analysis, could lead to two apparently different difficulties, both of which occur in the SLI population: delayed "first use" (competence) and delayed consistent use (performance) of an inflectional morpheme. In order to learn an inflectional system, the child must first analyse words into morphemes. Once the child has analysed the morphological elements and has learned the relevant grammatical system, she has attained competence with that particular grammatical structure. Without adequate morphological analysis skills, the attainment of competence could be delayed. Grammatical competence, however, does not lead immediately (if ever) to perfection in performance. In order to produce the morpheme in question correctly 100% of the time, the child must monitor her output, note when she has made an error, and correct the error (see Bowey, 1988; Clark, 1978; Marshall & Morton, 1978 for examples of young children's spontaneous repairs and arguments that such repairs involve linguistic awareness/analysis). This is the second role of linguistic analysis. A deficit in morphological analysis would make the attainment of consistently correct morphological performance more difficult.

These two roles of linguistic analysis both require the analysis of words into morphemes; first, as an automatic process of language acquisition, then as an on-line means of comparing productions to the internal grammar to check for accuracy. These skills can be considered primary linguistic activities, in the sense of Mattingly (1972). Such skills gradually become available to conscious introspection, providing the child with more and more explicit insights into grammatical structure. These same skills that allow the child to analyse linguistic input and monitor her own production can be applied to the speech of others, leading to more overt, more *meta-* linguistic analysis. Such overt analysis abilities develop into the skills necessary to do tasks less directly related to primary linguistic activities which can then be applied to secondary activities such as reading and writing and, arguably, experimental tasks. The application of linguistic analysis skills to secondary tasks might be fostered by exposure to and instruction in such tasks, as in, for example, reading and writing instruction.

Why might a child have difficulty in morphological analysis? Morphological systems are clearly specific to particular languages. While some linguistic properties might indicate generally what type of morphological system exists in a language, the actual items must be learned by the child. It is difficult to imagine linguistic universals that would guide language-specific morphological analysis; no general linguistic principle will tell a child to look for final /s/ in English as a morphological marker. In contrast, it has been proposed that innate universal principles do guide the acquisition of syntax (Chomsky, 1981). Morphological analysis of linguistic input might thus be more difficult than syntactic analysis guided by the principles and parameters of a Universal Grammar (such as outlined by Chomsky, 1981, for example). Thus, SLI children with linguistic analysis difficulties might be expected to have difficulty with the acquisition of idiosyncratic language-specific information, information that is stored in the lexicon.

Thus, it is hypothesized, first, that SLI children have more difficulty in the acquisition of language-specific information than with the acquisition of structures subject to universal linguistic principles; and second, that the difficulty with language specific structures is due to a deficit in linguistic analysis skills. If this is true, then SLI children should demonstrate normal facility in the acquisition of a structure subject to universal principles but demonstrate deficits in tasks requiring analysis of morphological structure.

Question 1: Development of Universal and Language-Specific Structures

In order to address question (1) and explore more fully the possible difference between the acquisition of structures involving innate universal principles (e.g., syntax) and the acquisition of more language-specific properties (e.g., morphology), an investigation of the acquisition of a grammatical structure with both complex syntax and complex morphology would be helpful. The passive structure in English meets this requirement.

In the principles and parameters framework (Chomsky, 1981), the syntax of the passive requires knowledge of the universal principles of case theory, theta-theory and the formation of argument chains (A-chains) (see Baker, Johnston, & Roberts, 1989; Borer & Wexler, 1987 for detailed analyses). It will be assumed here that the subject noun phrase originates in object position, where it receives a theta-role, which identifies which grammatical relation it plays in the sentence. The noun phrase also needs case, but cannot receive it in object position (because of the presence of the passive morphology, which is said to absorb case). As a result, it must move to subject position where it can receive case, thereby forming a passive sentence. Thus, in order to produce a passive sentence, the child must know the requirements of case assignment, theta-role assignment and be able to move noun phrases from one argument position to another (argument- or A-movement).

Passives can be formed with *get* as well as *be*. While it has been argued that *get* passives have a different syntactic structure than *be* passives, *get* passives still require the knowledge of theta-theory, case theory and A-chains (Fox & Grodzinsky, 1992; Haegeman, 1985; Hoshi, 1991; Lasnik & Fiengo, 1974) and as such, are of interest in this study.

Passive constructions can be either verbal or adjectival in nature. It is the verbal, not the adjectival form of the passive which is of interest in this study, since only the verbal passive requires the syntactic operation of A-movement (Borer & Wexler, 1987; Wasow 1977). The presence of a by-phrase is one indicator of a verbal rather than an adjectival passive. However, verbal passives may have, but do not have to have, a by-phrase.

The morphological complexity of the passive involves the multiple forms of the passive inflection (*ed* or *en)* and possible vowel changes in the stem (e.g., *bite-bitten*).

There has been some debate as to young children's ability to produce passives. Truncated passives (i.e., passives without by-phrases) have been noted to occur more frequently than full passives (i.e., those with by-phrases) in the elicited and spontaneous speech of young children (Baldie, 1976, Horgan, 1977), leading some to claim that full verbal passives are not produced by young children (Borer & Wexler, 1987). However, other researchers report full passives produced by 3 to 5 year-olds in elicited production tasks (Crain, Thornton and Murasugi (1987) and Crain and Fodor (1993)).

The exploration of the passive in SLI children has also indicated difficulty with the structure. The literature reveals few examples of passives in the speech of SLI children. Leonard (1989) suggests that this is not due to the low frequency of occurrence of passives, given that they do appear in the speech of normal children at an early age (Pinker, Lebeaux, & Frost, 1987). Menyuk and Looney (1972) found that SLI children performed more poorly on the repetition of passive sentences than a group of normal children matched on receptive vocabulary and tended to omit grammatical morphemes such as *is* and *by* in their repetitions.

Given the results of the above studies, an elicited production paradigm is the most appropriate technique for this study. It is most practical to study the child's expression rather than comprehension, since it would be difficult to differentiate between the comprehension of the passive morphology versus the passive syntax. Elicited production avoids the difficulty of the low frequency of passive constructions in spontaneous speech and allows for the collection of an adequate amount of data for analysis. Furthermore, it reduces the difficulty of distinguishing between the verbal or adjectival nature of the children's productions. In order to be confident that the children are producing true verbal passives, full passives with by-phrases should be elicited whenever possible. In the event that by-phrases are not always elicited, a carefully constructed elicitation protocol will aid in the analysis. A truncated passive can be interpreted as a verbal passive if it is produced in response to a situation in which a verbal passive and not an adjectival passive is the appropriate response. The proposal that SLI children suffer deficits only in the acquisition of language-specific information will be supported if the SLI children demonstrate proficiency with the syntax of verbal passives, implying the presence of a syntactic form of the passive inflection, while they continue to have difficulty with the morphological properties of the passive.

Question 2: Development of Linguistic Analysis Skills

In order to address question (2) and investigate the hypothesis that SLI children suffer from a deficit in linguistic (morphological) analysis skills a thorough investigation of morphological analysis tasks with a range of difficulty is required. Previous research has indicated that normal children develop linguistic analysis skills at a young age and that these continue to develop as the child grows older (Clark, 1978). Normal children have been shown to be able to analyse phonological and morphological structure in grammaticality judgment tasks as early as 3 to 5 years of age (Smith-Lock & Rubin, 1993). SLI children have shown varying success, performing the same as language-matched peers in some studies (Rubin, Kantor, & Macnab, 1990) and differently from language-matched peers in others (Kamhi & Koenig, 1985).

Standard metalinguistic analysis tasks, such as the judgment task, require explicit understanding of linguistic form. However, tasks with less explicit analysis requirements must be developed in order to tap skills that are more closely related to the analysis required in the initial learning of inflectional systems. The linguistic analysis associated with primary language acquisition appears to occur in a very automatic fashion. Thus, tasks which allow the child to use the primary language system automatically should be the easiest. Tasks should increase in difficulty to the extent that they require explicit analysis of the primary linguistic system.

The Normal Control Group: Language Matching

The syntactic and morphological skills of normal and SLI children should be compared in groups matched for language abilities. While a difference in performance between SLI and age-matched peers would be of interest, indicating that linguistic analysis skills are tied to expressive language ability rather than non-linguistic cognitive development, the comparison of most interest is SLI versus normal children of the same language level. Only by comparing language-matched groups can it be determined whether the SLI children have a deficit in morphological analysis abilities over and above what would be expected on the basis of their primary language deficit. As well, language-matching will allow for the comparison of the development of various components of the grammar in children matched on one of the components.

The method of language matching is critical to the study. Matching on the basis of expressive rather than receptive language seems most appropriate, since the ability to manipulate morphological structure consciously would likely require expressive knowledge of the structure. The children should be matched on their spontaneous speech, since formal testing removes the child from the realm of spontaneous and automatic output, and therefore, might introduce linguistic analysis skills into the task. Mean length of utterance (MLU) is one possible measure of language development using spontaneous speech. However, MLU does not provide information regarding what type of structures are used, thus it is not possible to distinguish between an MLU based on grammatically simple but lengthy utterances and one based on grammatically complex utterances. Therefore, the possibility of matching children with different linguistic skills is significant. Furthermore, the correlation of MLU with grammatical development decreases in the later stages of language acquisition (Scarborough, Rescorla, Tager-Flusberg, Fowler, & Sudhalter, 1991). Thus, MLU is not the most appropriate matching technique.

Given the relatively consistent order of acquisition of grammatical morphemes noted by Brown (1973), children who have acquired a particular morpheme can be assumed to have attained the same level of grammatical (morphological) development. Thus, if only children who use the regular past tense (*-ed)* consistently are included in the study, the subjects will have acquired most other inflectional morphemes. This will establish a minimum level of development. Further, all children go through a stage in which they overgeneralize regular endings to irregular stems (as in *goed* for *went).* This stage coincides with the use of the regular *-ed* form (Marcus, Pinker, Ullman, Hollander, Rosen, & Xu, 1992). If only children who are in the stage of overgeneralization are included, a minimum and maximum level of grammatical development will be established. All subjects will have acquired the regular past tense, but they still will not have acquired the irregular past tense forms. In this way, subjects can be matched on expressive language without reliance on MLU and without the confounds of linguistic analysis abilities required by formal testing.

EXPERIMENT 1

In order to address the experimental questions of whether SLI children show normal facility in syntax and a deficit in morphology, and whether their morphological deficits could be attributed to poor morphological analysis skills, language-matched groups of normal and specifically-language-impaired children were compared in the following study.

Method

Subjects. Sixteen normal and seventeen specifically language-impaired (SLI) children were included in the study. All of the children had normal vision and no known hearing loss, were monolingual speakers of English, demonstrated non-verbal intelligence within the average range on the Block Design and Geometric Design subtests of the *Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R)* (Wechsler, 1989) and met the language and screening criteria outlined below. All of the children were attending preschool or elementary school in Southern Ontario, Canada. The SLI children had been previously identified as SLI in their elementary schools by certified speech-language pathologists. The SLI children ranged in age from 5;4 to 7;3, with a mean age of 6;2. The language-matched group ranged in age from 3;3 to 4;3, with a mean age of 4;0.

Language screening

Ten verbs, which in the adult language have irregular past tense forms, and six verbs with regular past tense forms (two verbs for each allomorph of the past tense morpheme) were elicited from the children in a story telling task. Stories were acted out with the child using toys. The child was then asked to tell the experimenter what had happened so the experimenter could write the story down, thus eliciting the past tense. Children who had not yet acquired the correct irregular form of at least five out of ten of the irregular verbs, but who did use the /d/ and /t/ allomorphs on the regular verbs, were included in the study. The screening stimuli can be found in Appendix A.

Articulation screening

Children were asked to repeat words containing final /s/, /z/, /t/ and /d/ which were not inflectional morphemes (e.g., *act, collapse*). All of the final consonant clusters found in the experimental tasks were included in this task. Real words were used wherever possible. Because the addition of morphemes sometimes creates consonant clusters which would not otherwise be permitted, it was not always possible to use real words. In such cases, non-words were used. Only children who could produce these consonant clusters were included in the study. This ensured that any omissions of inflectional morphemes in the experimental tasks were due to the nature of the task and not to articulatory difficulties.

Words containing later-developing speech sounds such as /tʃ/, /l/ and /ɹ/ were also elicited in order to establish the current articulatory pattern of the child. No children were excluded on this basis. However, the information was considered in the scoring of the experimental tasks, so that children would not be erroneously assumed to be making explicit changes in sound structure when they were actually making a developmental articulation error.

Subject referral and selection

Normal children were selected from those children for whom parental permission was obtained and who fell roughly within the age range of 3;6 to 4;6. A total of 37 children were screened for the LM group. 21 did not meet the language screening criteria.

Referrals of SLI children were obtained by asking school speech-language pathologists to refer children who met the following criteria: specifically language impaired, normal non-verbal skills, monolingual English speakers, no history of hearing loss, 6 - 7 years old, speech intelligible enough for reliable data collection. These criteria were used as a *guideline only*. The speech-language pathologists were encouraged to refer anyone they thought might be appropriate. A total of 76 SLI children were referred. 17 of those were included in the study. Of those who were excluded, 34 did not meet the language screening criteria, 11 had a history of hearing loss (including fluctuating conductive loss), eight did not pass the articulation screening, four scored below average on the assessment of non-verbal performance (*WPPSI-R*) and one had no available non-verbal intelligence information. 85% of the children excluded on the basis of the language screening used overgeneralizations in the screening task. The remaining 15% (5 children) used the correct irregular forms, as expected for their age. Only one child who met the screening criteria on irregular verbs was dropped from the study due to inconsistent use of the regular past tense.[2] The subjects' performance on the screening task can be seen in Appendix B.

Experimental tasks: Real word sentence completion

In this task, the child was told that the experimenter would start a story and that the child was to finish it, with just one word about the picture. For example, the child was shown a picture of woman at a grocery store with a cart full of groceries. The experimenter stated, "This woman is shopping. Every day, she _____." The child was expected to respond "shops." Two training trials were provided, with feedback. No feedback was provided to experimental trials.

The stimulus sentences required the manipulation of the morphemes for regular past tense, third person singular present tense, and the present progressive tense. Each of these inflectional morphemes occurred in five stimulus sentences and five responses. For the past and present tense morphemes, the stimuli and responses contained two instances each of the voiced and voiceless allomorphs and one instance of the shwa + consonant allomorph.

This task was intended to be very similar to spontaneous speech, with only a minimal reduction of automaticity, since the addition of the inflection should be fairly automatic, given the correct stem. However, the task involved some morphological analysis in that it required the subject to analyse the verb into stem + inflection and to replace one inflection with another. This task could be distinguished from spontaneous speech in that the child had to complete a sentence with a particular single word and perform the appropriate morphological manipulation, thus going beyond the automatic nature of spontaneous speech.

Non-word sentence completion

This task was similar to task (1) except that nonsense words were used instead of real words (e.g., "This guy *linged* yesterday. Every day he _____"). The child was provided with the following instructions. "These pictures are just like the first ones. I'll start a story and you finish it. The only difference is that these are silly pictures, with silly names you probably haven't heard before." No training trials were included in this task. Instead, if the child responded with a word other than the nonsense word, she was told "You use the same word I use. So if I use *sput*, you use *sput* too." The stimulus sentence was not readministered following the cue. The same morphemes and allomorphs were used, with the same frequency as task (1).

This task was believed to require slightly more morphological analysis than the real-word sentence completion. The child had to apply her morphological knowledge to a word she had not encountered before, further increasing the skills needed in addition to those required for spontaneous speech.

Comprehension of inflected non-words.

In this task, the children were shown a page divided into two sections. One section contained the picture of a novel item. The other section contained two of the same item. The task was introduced as follows. "I'm going to show you some funny pictures with some funny names. All you need to do is listen carefully and point to the picture I tell you to. OK?" With each new page, the experimenter said the following, changing the name of each nonsense item as appropriate. "This page has *pashes* on it. There are two in this part [pointing] and one in this part [pointing]. Point to the part that has the *pash*." The child then had to choose one of the sections of the page.

Six training trials were provided, which consisted of three nonsense items, with both the plural and singular tested. All of the training stimuli took the /əz/ allomorph because it was believed that its syllable status might make it the easiest. Feedback followed the training trials but not the experimental trials. The experimental trials consisted of ten nonsense items. Each was tested in the singular and the plural form, for a total of 20 test items. All three allomorphs were tested.

This task required morphological analysis in order to analyse a non-word into morphemes and explicitly understand that the /s/ ending marked plural. It can be distinguished from spontaneous speech because the child had only morphological information on which to base her response. The words were all unknown and no contextual information was available to cue the child, unlike ordinary conversation.

Judgment and correction of morphological errors

This task involved the use of a puppet who made morphological errors in his speech. Children were asked to judge, identify and repair these errors. A semantic judgment task was used as an introduction in order to familiarize the child with judgment tasks. This task also offered a means of highlighting the distinction between semantic and morphological judgments so as to reduce the likelihood that the children would make semantic judgments in the morphological task.

In the semantic task, the child was told that Ernie was a funny puppet and that he said silly things, things that just weren't true. Examples were provided in which the puppet called the experimenter by the wrong name and the experimenter identified the error and corrected the puppet. The puppet then called the child by the wrong name and the child was invited to correct the puppet. The child and the experimenter then acted out a story, agreed on a verbal description of what had happened, then asked the puppet to comment. The puppet's comment involved the substitution of an object noun (e.g., "Barbie ate a cookie" for "Barbie ate a pizza"), a subject noun (e.g., "The man went for a run" for "The lady went for a run") or a verb (e.g., "The man drank the french fries" for "The man ate the french fries") in a sentence. The same judgment, identification and repair protocol was used for the semantic and morphological tasks, and is outlined below.

In the morphological task, a different puppet, Bert, was introduced as a puppet who was not silly, unlike Ernie. It was explained that everything Bert said was true but that he said things the wrong way sometimes and that he wanted help to say things the right way. The child and the experimenter then acted out a story, agreed on a verbal description of what had happened, then asked the puppet to comment. 50% of the puppet's comments were grammatically correct and 50% involved the omission of an inflectional morpheme (e.g., "The boy has lots of toy"). The child was then asked

(1) Did he say it the right way or the wrong way? (judgment)
(2) What was the wrong part? (identification)
(3) Can you fix it? (repair)

After each trial, feedback was provided to the child. If the child responded correctly to all three questions she was told that she was right. If the child made an error on any of the three questions, the correct answer was explained. The item was repeated until the child responded correctly to all three questions, to a maximum of three trials. An example of the protocol can be seen below.

(i) story: Barbie eats 2 cookies.
(ii) experimenter (E) to the child (C): what did Barbie eat?
C: 2 cookies
(iii) E to the puppet: Bert, what did Barbie eat?
(iv) puppet: 2 cookie
(v) E to C: was that right or wrong?
(vi) C: right
(vii) E to C: I think it was wrong because he said 2 cookie instead of 2 cookie**s** (emphasis on /s/). She ate two, so he should have said cookie**s**, not cookie.
(viii) repeat to a total of three times, if necessary

The errors consisted of the omission of plural, possessive or past tense morphemes. For each of these inflectional morphemes, two phrases and one full sentence were included, for a total of nine items with errors. Nine parallel constructions without errors were included. All of the stems taking inflections ended in vowels so that when they were inflected the word ended in a single consonant, the voiced allomorph ([z] or [d]). This was done in order to simplify the phonological demands of analysing consonant clusters. In addition, two verbs which the child overgeneralized in the language screening were included. These verbs varied for each child.

The morphological judgment task required much more than the automaticity of spontaneous speech. It required the child to examine an utterance and consider the appropriateness of the linguistic form outside of the communicative intent. It required explicit knowledge of the grammatical constructions involved and the conditions for their use.

Child-generated errors

This task was identical to the judgment tasks outlined above, except that the child was asked to be the puppet. In the semantic task, the child was told to talk silly like Ernie. In the morphological task, she was told to say things wrong, like Bert. The experimenter then acted out a story and commented on it, providing a phrase or sentence for the child to manipulate. In the semantic task, the child was asked to manipulate the sentence *The man walked home*. In the morphological task, the child was asked to make errors on two plural phrases, two possessive phrases and two past tense verbs.

This task had the highest morphological analysis demands. In the morphological task, the child had to explicitly understand the morphological structure of the word. She had to know exactly what a morpheme was and exactly how it was manipulated in the judgment task in order to be successful in this task.

Elicitation of passive sentences

This task was included to investigate the dissociation of the acquisition of morphology and syntax, in order to determine whether the syntactic components of the passive were acquired before the morphological components.

Passive sentences were elicited from the children using a story-telling task, similar to the elicited production technique used by Crain, Thornton, and Murasugi (1987). A story, in which two agents acted upon two patients, was acted out with toys (e.g., a dog chased a pig and a cat chased a horse). The child was asked what happened to one of the characters in the story (e.g., "what happened to the pig?", "what happened to the horse?"). The expected response was a passive structure (e.g., "the pig was chased by the dog", "the horse was chased by the cat"). Ten such stories were used, each with two different passive-eliciting questions. Passives were elicited for the following verbs: lick, bite, fly, ride, eat, take, chase, drive, chop, throw. Prompting was sometimes necessary to elicit the passive. In such a case, the experimenter started the sentence with the passive subject and then stopped (e.g., "What happened to the pig? The pig..."). This strategy indicated to the child that she was to start the sentence with the passive subject and was frequently, but not always, successful in eliciting a passive construction.

Procedures. Each child was tested individually in a quiet room in their preschool or elementary school. They were seen for a total of three or four sessions approximately 30 to 45 minutes in length. The language screening was administered first. At each session, an attempt was made to elicit the passive. If no passive structures were elicited with the first three items, the task was discontinued, other tasks administered and the task was then attempted again at the next session. If no passives had been elicited after three sessions, no further attempts were made. The remaining tasks were administered in varying order (depending on the time available) except that the picture tasks were always administered in one session, in the order real word expression, comprehension, non-word expression. All of the tasks, with the exception of the comprehension task, were recorded on audiotape and later transcribed.

Results

Sentence completion tasks

Responses in both the sentence completion tasks were scored as correct or incorrect. In order to be considered correct, the response had to contain both the correct verb and the correct inflection. Use of a different verb with the correct inflection was considered an error in this scoring system. Scoring the data by crediting all correct inflections, regardless of verb, improved scores in both groups, but the relationship between the groups remained the same. Therefore, the original scoring system was maintained. Incorrect responses were further classified as omission of the correct inflection, a repetition of the inflection used in the stimulus sentence, or as another incorrect inflection. In the real word task, the mean score for the SLI group was 6.48 out of 15 (S = 3.11), and for the LM group, 6.69 (S = 3.03). Performance on the non-word task was lower: 4.29 out of 15 (S = 3.04) for the SLI group and 4.56 (S = 3.31) for the LM group.

A two-way analysis of variance with one between groups factor (diagnosis: SLI and language-matched (LM)) and one repeated measure (task: real word, non-word) showed no significant difference between the groups ($F < 1$), a significant difference between tasks ($F(1,31) = 28.49, p < .001$) and no interaction ($F < 1$). Thus, the SLI group performed the same as their language-matched peers. The real word sentence completion task was significantly easier than the non-word task.

The results of the error analysis can be seen in Table 1. A two-way analysis of variance with one between group factor (diagnosis: SLI, LM) and one repeated measure (task: real word, non-word) was performed for both repetition and omission errors. There was a significant difference in the number of repetition errors between the real and non-word tasks ($F(1,31) = 45.34, p < 0.001$) but no significant group difference ($F < 1$) and no significant interaction ($F(1,31) = 3.69, p > .05$). With omission errors, there was no significant task effect ($F(1,31) = 0.94, p > .05$), no significant group effect ($F(1,31) = 2.9, p > .05$) and no significant interaction ($F(1,31) = 0.01, p > .05$). Thus, the LM and SLI children made the same number and type of errors, with more repetition errors occurring in the non-word task than the real word task.

Table 1. *Real word and non-word sentence completion. Mean number of repetition and omission errors (standard deviation in brackets).*

	Repetition		Omission	
	real word	non-word	real word	non-word
LM group	2.5 (2.18)	5.4 (3.52)	1.57 (1.60)	1.2 (1.86)
SLI group	2.23 (2.31)	6.47 (3.43)	2.53 (2.85)	2.12 (1.87)

Comprehension of Inflected Non-Words

The comprehension task had a maximum score of 10. Since the response required a choice between two options, a score of five indicated chance performance. The LM group received a mean score of 5.69 ($S = 1.58$) and the SLI group, 6.29 ($S = 1.9$). A one-group t-test indicated that the performance of LM group did not differ significantly from chance ($t(15) = 1.74$, $p > .05$) while the performance of the SLI children did ($t(16) = 2.81$, $p < .05$). Nevertheless, a comparison of the two groups showed no significant difference in performance between the SLI and LM children ($t(31) = -0.10$, $p > .05$).

Subjective data from the test administration indicated that performance on this task was "all or none." In other words, the children either figured it out or they guessed. Those children who figured it out generally did so during the training sessions and often spontaneously commented on their discovery of how to do the task (e.g., "I heard you say *pash* and that's one *pash*"). When asked afterwards how they decided the right answer, some of the children who had done well explained that the examiner had told them to point to one or two (e.g., "I just heard 2-2-1-2"), while most of the unsuccessful children said they had guessed, or alternated between the top and bottom picture.

If those children who received a score of 8 or higher are considered to have understood the task, (the majority of children received a score within 2 points of the chance score (5 ± 2)), one child in the LM group (SW, 4;0) and four children in the SLI group (MQ, 5;10, kindergarten; MD, 6:8, grade 1; BE, 6;5, grade 1 and TK, 6;9, grade 1) could do the task. It is interesting to note that three of the four SLI children who could do the task were in grade one and, therefore, had had reading and writing instruction.

Judgment Task

Children received a score on the basis of the number of incorrect stimulus items identified as incorrect. The nine test items yielded a maximum score of nine for each of judgments, identifications and repairs, for each of the three trials. Scoring was cumulative, so that if a child scored correctly on trial 1 and therefore did not receive trials 2 and 3, she received credit for the correct response in the score of trials 2 and 3. Thus, a score of 7 out of 9 correct judgments on trial 3 indicates that, by trial 3, the child had made 7 correct judgments. She may have responded correctly to 2 items on trial 1 (yielding a trial 1 score of 2), 3 items on trial 2 (yielding a trial 2 score of 5) and 2 items on trial 3 (yielding a trial 3 score of 7). In order to preserve this type of information each trial was analysed separately, rather than examining only trial 3, or creating a composite score based on all 3 trials. Judgments of correct items were not included.

A correct judgment was considered a response of "wrong" to the question "Did Bert say it right or wrong?." A correct identification was considered the repetition of the entire phrase or sentence, with the error, or the repetition of the erroneous word. A correct repair was considered the repetition of the erroneous word, phrase or sentence, with the error corrected. An example of a typical response can be seen below.

stimulus: "The lady dress is white."
judgment: wrong
identification: "The lady dress is white" or "the lady dress."
repair: "The lady's dress is white" or "the lady's dress."

The results can be seen in Table 2 and Figure 1. In order to compare the performance of the SLI and LM groups, a two-way analysis of variance was performed, with one between-groups factor (diagnosis: SLI, LM) and one repeated measure (task: judgment, identification, repair). There was no significant effect for group on any of the trials (trial 1: $F < 1$; trial 2: $F < 1$; trial 3: $F < 1$). There was a significant effect of task, for all three trials (trial 1: $F(2,31) = 47.5$, $p < 0.001$; trial 2: $F(2,31) = 30.94$, $p < 0.001$; trial 3: $F(2,31) = 32.54$, $p < 0.001$). There were no interactions (trial 1: $F < 1$; trial 2: $F < 1$; trial 3: $F(2,62) = 1.79$, $p > .05$). Thus, the SLI group performed in the same way as their language-matched peers.

Table 2. *Judgment task. Mean correct (out of 9) (standard deviation in brackets).*

	judgment		identification		repair	
	SLI	LM	SLI	LM	SLI	LM
trial 1	5.29 (1.53)	5.88 (1.86)	2.94 (2.08)	3.00 (2.53)	3.77 (1.92)	4.00 (1.86)
trial 2	7.65 (1.41)	7.32 (1.62)	4.65 (2.74)	4.38 (3.05)	6.41 (2.0)	5.38 (2.66)
trial 3	8.12 (1.22)	8.19 (1.38)	5.41 (2.53)	5.25 (3.15)	7.24 (2.08)	6.06 (2.29)

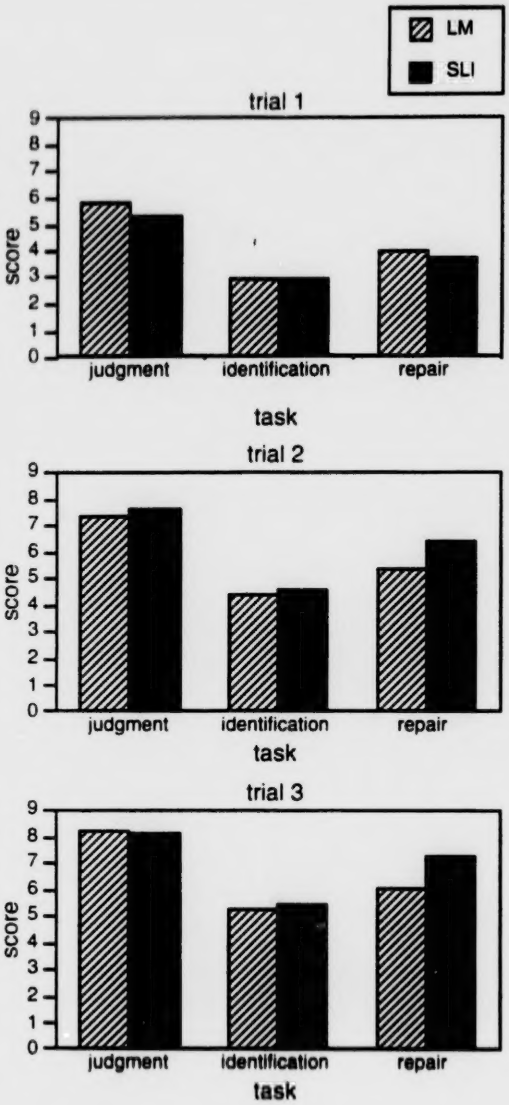


Figure 1. **Judgment task.**

To determine the extent of improvement over the three trials, a two-way analysis of variance was performed, with one between-groups factor (diagnosis) and one within-groups factor (trial), for judgments, identifications and repairs. There was a significant improvement over trials for judgments ($F(2,31) = 63.15$, $p < 0.001$), identifications ($F(2,31) = 60.85$, $p < 0.001$) and repairs ($F(2,31) = 66.19$, $p < 0.001$). There were no significant interactions between group and trial for judgments ($F(2,62) = 1.86$, $p > .05$) or identifications ($F < 1$). There was a significant interaction for repairs ($F(2,62) = 4.80$, $p < 0.01$). Thus, both the normal and the SLI groups improved over trials. For repairs, the SLI group improved more over trials than the normal group.

The children had two opportunities to judge their own overgeneralizations. Examination of their responses indicated that children almost always accepted their own production as correct. 92% of overgeneralizations were accepted by the LM group and 82% were accepted by the SLI group. No one was able to correct her own overgeneralization error.

Child-generated errors

A correct response to this task required the child to omit the inflectional morpheme from the stimulus item provided. For example, the correct response to "two eyes" was "two eye." Phonological changes ("two byes") and semantic changes ("one eye") were considered incorrect. A child received one point for each correct response, for a possible total of 6. The SLI group received a mean score of 1.76 (S = 2.28). The LM group received a mean score of 0.69 (S = 1.35). The groups' performance did not differ significantly ($t(31) = -1.64$, $p > .05$). This task was quite difficult for all the children. A qualitative analysis of the responses indicated that 57% of the LM and 34% of the SLI responses involved no change to the stimulus item, 24% of the LM and 22% of the SLI group's responses were semantic changes, 6% of the LM and 14% of the SLI responses were phonological changes and 12% of the LM and 31% of the SLI responses were morphological changes. Thus, although non-significant, differences do exist, with the SLI children being more able to create morphological changes than the younger LM children. This may reflect the analytic ability gained through reading and writing instruction that the SLI, but not the LM children, have received (due to further years of schooling).

Elicitation of Passive Sentences

Passive constructions were elicited from the majority of the children in the study. All of the

SLI children produced passives, with a mean of 15 per child. 59% of these children (10 out of 17) produced full passives with prepositional phrases. In the LM group, 12 out of 16 children produced passives, with a mean of 11 per child. 42% (5 out of 12) of the children produced full passives. Thus, both groups were able to productively generate syntactically correct passive structures. No syntactic errors were noted. If children failed to produce a passive sentence, they produced an active equivalent. Almost all of the passives elicited were got-passives, although some be-passives were elicited. Examples of the children's productions can be seen below.

(1) he got licked by a tiger (MK, 7;3, SLI)
(2) it got taken by the man (AG, 6;11, SLI)
(3) it got eaten by the big horse (SW, 4;1, LM)
(4) it got pushed down by the girl (MW, 4;2, LM)
(5) it's gonna be ride (JC, 5;4, SLI)
(6) (the fries) was eated (JP, 4;1, LM)
(7) it got licked by the horse (AG, 6.11, SLI)
(8) it got chased by the dog (AG, 6.11, SLI)
(9) he got chopped off (AM, 3;11, LM)
(10) the two babies got licked (KM, 3;11, LM)

Prepositional errors occurred in both groups. *from* was substituted for *by* in 23 cases (28%) in the SLI group and 3 cases (9%) in the normal group. *with* was substituted for *by* in 3 cases in the SLI group.

(11) the tree got knocked over from the baby (IP, 6;0, SLI)
(12) he got eaten from Mickey (AM, 3,11, LM)
(13) he got licked with the pig (AP, 7;2, SLI)

Morphological errors were common in both groups. The errors took a variety of forms including the incorrect use of *-ed, -en,* both or neither, combined with either a present or past tense stem. The use of the present or irregular past form as the stem was not associated with whether or not the correct form contained a vowel change. Examples of the error types can be seen in Table 3. The frequency of each type of morphological response can be found in Table 4. All tokens of the passive were included in this calculation, including repeated productions. The SLI and LM groups, for the most part, used the various morphological forms with similar frequency. However, the LM children tended to use forms with the *past* stem more often than the SLI children.

Table 3. *Morphological error types in passive elicitation.*

stem only:
    (1) he got chase around (JG, 5;9, SLI)
    (2) it got ride by the baby (MW, 4;2, LM)
stem + *ed* :
    (3) he got bited from the horse (MD, 6;8, SLI)
    (4) it got eated (SW, 4;0, LM)
stem + *ed* + *ed* :
    (5) it got throweded and this one got throweded (JM, 6;7, SLI)
    (6) the car got driveded (CG, 3;11, LM)
stem + *en* :
    (7) (he) got chasen by that (MW, 4;2, LM)
    (8) it got drive-en (BE, 6;5, SLI)
stem + *ed* + *en* :
    (9) he got chaseden (BE, 6;5, SLI)
    (10) it got throweden (BE, 6;5, SLI)
stem + *en* + *ed* :
    (11) the hotdog got eatened up (IP, 6;0, SLI)
    (12) it got eatened (MW, 4;2, LM)
past stem:
    (13) (the ball) gotted took (MK, 7;3, SLI)
    (14) it got rode on (SW, 4;0, LM)
past + *ed* :
    (15) it got ated from the boy (HB, 5;4, SLI)
    (16) it got tooked (DM, 4;0, LM)
past + *ed* + *ed* :
    (17) the car got stoleded (MQ, 5;10, SLI)
past + *en* :
    (18) it got droven from Mickey Mouse (HB, 5;4, SLI)
    (19) both of them got aten up (KH, 4;2, LM)
past + *ed* + *en* :
    (20) the ball got stoleden (MQ, 5;10, SLI)
    (21) it got stoleden (KH, 4;2, LM)
past + *en* + *ed* :
    (22) it got tookened from the boy (HB, 5;4, SLI)
    (23) it got atened (DM, 4;0, LM)

Table 4. *Frequency of morphological responses (%) in passive elicitation.*

group	correct	stem +∅	stem +ed	stem +en	past +∅	past +ed	past +en	other
LM	39.49	9.24	22.69	0.08	2.52	5.88	12.61	6.72
SLI	41.92	10.00	26.92	3.00	1.15	3.85	2.31	10.39

Individual subject data demonstrated patterns of *-ed* and *-en* usage. The children could be classified as predominantly *ed* -users, predominantly *en* -users, or mixed *-ed* and *-en*. A child was considered a mixed *-ed* and *en*-user if she used both endings more than once in the task. A child was still considered an *en*-user if she produced the regular *ed* verbs correctly. In the SLI group, 9 children were *ed* -users and 5 children were mixed. That is, children either used *-ed* for all *en* verbs, or used a mixture of *-ed* and *-en*. None used *-en* on all the verbs requiring it. 3 children did not provide enough data for analysis. In the LM group, 4 children were *ed* -users and 3 children were mixed. Again, no children always used *-en* when appropriate. 8 children did not provide enough data for analysis. While the children used most regular *-ed* forms correctly, two of the 'mixed' children (one SLI and one LM) used *-en* in place of the correct *-ed*. Thus, both overgeneralization of *-ed* to *-en* verbs and overgeneralization of *-en* to *-ed* verbs occurred. In cases where both *-en* and *-ed* were added to a stem, they were not always added in the same order. Both *-eden* and *-ened* were produced by some children. Examples of each pattern can be seen in Table 5.

Table 5. *Patterns of passive morpheme use.*

ed -user (CG, 3;11, LM)

(1) it got bite
(2) you got licked
(3) it got atened
(4) it got tooked
(5) it got throwed
(6) it got droved
(7) it got knocked down
(8) it got rided
(9) it got chased

Mixed (BE, 6;5, SLI)

(1) he got licked
(2) he got lick
(3) he got bited
(4) he got flied over
(5) he got squished
(6) it got ated…eaten
(7) it got eaten too
(8) he got chaseden
(9) he got riden too
(10) he got takeden
(11) it got taken too
(12) it got drivened (drive+ened)
(13) it got driven (drive+en) too
(14) he got knocked down
(15) it got throweden

Summary of Results

The SLI and LM groups were both capable of producing passive syntax without error, but made many errors with passive morphology. The groups did not differ significantly on the morphological analysis tasks.

EXPERIMENT 2

The results of the first experiment indicated no difference in the performance of the SLI and language-matched normal groups. In order to compare the performance of the SLI children with age-matched peers and to confirm that they were performing at a lower level than might be expected for their age, a second experiment was conducted.

Method

Subjects. Sixteen normal children were included in this experiment. They ranged in age from 5;7 to 6;5, with a mean age of 6;0. The children met all the same criteria outlined for the subjects in the first experiment , with the exception that only children who overgeneralized on less than 5 out of 10 of the verbs in the language screening were included. The children in this study did not differ significantly in age ($t(31) = 0.872, p > .05$) from the SLI group in Experiment 1.

Tasks and procedures. The same tasks and procedures were used in this study as were outlined for Experiment 1.

Results

Sentence completion tasks

The same scoring procedure was used as was outlined for Experiment 1. The AM group received a score of 10. 25 out of 15 correct ($S = 2.21$) on the real word task and 6.5 ($S = 2.56$) on the non-word task, compared to the SLI performance of 6.48 ($S = 3.11$) on the real word task and 4.29 ($S = 3.04$) on the non-word task. A two-way analysis of variance with one between groups factor (diagnosis: SLI and age-matched (AM)) and one repeated measure (task: real word, non-word) indicated a significant difference in performance between the SLI and age-matched (AM) groups ($F(1, 31) = 12.30, p < .001$), a significant difference in performance on words versus non-words, ($F(1, 31) = 43.87, p < .001$) and no significant interaction ($F(1,31) = 3.14, p > .05$). Thus, the SLI group performed significantly worse than their age-matched peers. The real word sentence completion task was significantly easier than the non-word task.

The results of the error analysis can be found in Table 6. In order to compare the number of repetition errors made by each group in the real word and non-word tasks, a two-way analysis of variance with one between groups factor (diagnosis: AM, SLI) and one within groups factor (task: real word, non-word) was performed. There was a significant effect for task ($F(1,31) = 97.48$, $p < 0.001$), but no effect for group ($F(1,31) = 0.45$, $p > .05$) and no significant interaction ($F(1,31) = 1.61, p > .05$). A similar analysis of the omission errors found a significant effect for group ($F(1,31) = 9.58, p < 0.01$), but no effect for task ($F < 1$) and no significant interaction ($F(1,31) = 1.97, p > .05$). Thus, the SLI children made the same number of repetition errors, but significantly more omission errors than their age-matched peers. There was no difference in the number of omission errors between the non-word and real word tasks. However, more repetition errors were made in the non-word task.

Table 6. *Experiment 2: Real word and non-word sentence completion. Mean number of repetition and omission errors (standard deviation in brackets).*

	Repetition		Omission	
	real word	non-word	real word	non-word
AM group	2.56	5.88	0.38	1.19
	(1.55)	(2.25)	(0.72)	(1.42)
SLI group	2.23	6.47	2.53	2.12
	(2.31)	(3.43)	(2.85)	(1.87)

Comprehension of Inflected Non-Words

As with Experiment 1, the comprehension task had a maximum score of 10. Since the response required a choice between two options, a score of five indicated chance performance. The AM group received a mean score of 7.56 (S = 2.03). A one-group *t*-test indicated that the performance of this group differed significantly from chance ($t(15) = 5.04, p < 0.001$). The mean score for the SLI group was 6.29 (S=1.9). A comparison of the SLI and the AM groups showed no significant difference in performance ($t(31) = -1.86, p > .05$). Nevertheless, 8 children in the AM group (compared to 4 in the SLI group) met the success criterion of eight correct responses. Some children provided interesting insight into the task through their spontaneous comments. For example, one child explained "*mooz* means one but *moozes* means two. So, you said *moozes*, so it's two. This is by numbers."

Judgment Task

The data were scored as in Experiment 1. The results can be found in Table 7 and are represented graphically in Figure 2. In order to compare the performance of the SLI and AM groups, a two-way analysis of variance was performed, with one between-groups factor (diagnosis: SLI, AM) and one repeated measure (task: judgment, identification, repair). The SLI group differed significantly from the AM group on all three trials (trial 1: $F(1,31) = 26.45, p < 0.001$; trial 2: $F(1,31) = 16.47, p < 0.001$; trial 3: $F(1,31) = 13.86$, $p < 0.001$). There was a significant effect of task on all three trials (trial 1: $F(2,31) = 31.78$, $p < 0.001$; trial 2: $F(2,31) = 28.71, p < 0.001$; trial 3: $F(2,31) = 24.35, p < 0.001$). There was no significant interaction for trial 1 ($F < 1$).

Table 7. *Experiment 2: Judgment task. (Mean correct (out of 9) (standard deviation in brackets).*

	judgment		identification		repair	
	SLI	AM	SLI	AM	SLI	AM
trial 1	5.29	7.25	2.94	5.5	3.77	6.19
	(1.53)	(0.93)	(2.08)	(1.37)	(1.92)	(1.12)
trial 2	7.65	8.69	4.65	7.38	6.41	8.38
	(1.41)	(0.48)	(2.74)	(1.5)	(2.0)	(0.5)
trial 3	8.12	8.94	5.41	8.06	7.24	8.75
	(1.22)	(0.25)	(2.53)	(1.34)	(2.08)	(0.45)

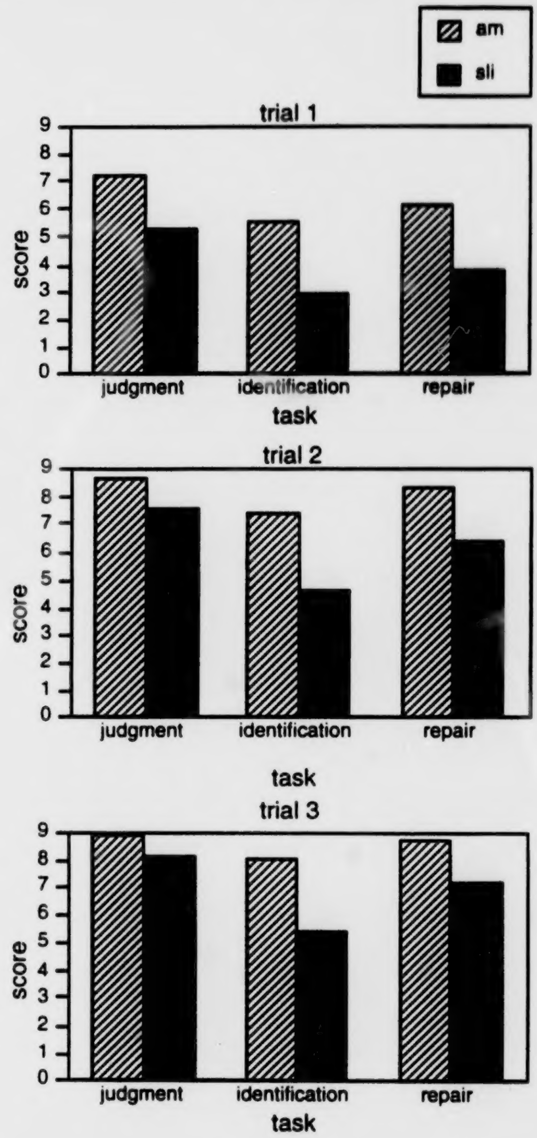


Figure 2. **Experiment 2: Judgment task.**

There were significant interactions for trials 2 and 3 (trial 2: $F(2,62) = 4.19$, $p < .05$; trial 3: $F(2,62) = 5.97$, $p < 0.01$). Thus, the SLI group performed worse than their age-matched peers on the judgment, identification and repair of errors, on all three trials. Furthermore, there was a difference in performance on judgments, identifications and repairs for both groups on trial 1, but only for the SLI group on trials 2 and 3.

To determine the extent of improvement over the three trials, a two-way analysis of variance was performed, with one between-groups factor (diagnosis) and one within-groups factor (trial), for judgments, identifications and repairs. There was a significant improvement over trials for judgments ($F(2,31) = 82.44$, $p < 0.001$), identifications ($F(2,31) = 87.18$, $p < 0.001$) and repairs ($F(2,31) = 166.63$, $p < 0.001$). There was a significant interaction between group and trial for judgments ($F(2,62) = 5.01$, $p < 0.01$) and repairs ($F(2,62) = 3.34$, $p < .05$). There was no significant interaction for identifications ($F < 1$). Thus, both the AM and SLI groups improved over trials. The AM group showed a ceiling effect for judgments and repairs.

Child-generated Errors

The AM group produced a mean of 1.6 ($S = 2.09$) self-generated errors, compared to 1.76 ($S = 2.28$) in the SLI group. This difference was not significant ($t(30) = -0.38$, $p > .05$). This task was quite difficult for all the children. A qualitative analysis of the responses in the AM group indicated that 35% of the responses involved no change to the stimulus item (compared to 34% in the SLI group), 26% were semantic changes (SLI group: 22%), 9% were phonological changes (SLI group: 14%) and 29% were morphological (SLI group: 31%). Thus, the AM and SLI groups performed similarly on this task.

Elicitation of passive sentences

All of the children in the age-matched group produced passive constructions, with a mean of 22 per child. 13 out of 15 (87%)[3] of the AM children produced full passives. No syntactic or prepositional errors were noted. Morphological errors were common. As in Experiment 1, errors took on a variety of forms consisting of the present or irregular past as a stem, plus *en*, *ed* or both, including the overgeneralization of *en*.

Examples of Correct Productions

(14) the pig got chased by the tiger (KG, 6;5)
(15) he got licked by the dog (RB, 6;4)

Table 8. *Frequency of morphological responses (%) in passive elicitation*

group	correct	stem +∅	stem +ed	stem +en	past +∅	past +ed	past +en	other
AM	42.08	0.08	15.78	14.52	9.16	1.25	10.79	5.39
SLI	41.92	10.00	26.92	3.00	1.15	3.85	2.31	10.39

Table 9. *Experiment 2: Patterns of passive morpheme use.*

ed-user (KF, 6;5)
(1) he got licked by the tiger
(2) he got bited by the horse
(3) he got flied over by the horse
(4) the cat got chased by the dog
(5) the dog got chased by the cat
(6) it got eated and the hotdog got eated
(7) he got catched
(8) he got chased by the man
(9) it got throwed
(10) it got drove
(11) it got chopped by a lady

mixed (RB, 6;4)
(12) he got licked by the dog
(13) he got bit by the tiger
(14) he got eaten up
(15) it got chased by the girl
(16) it got takened
(17) it got tooken too
(18) it got droven
(19) it got throwed
(20) it got cut down

en -user (RK, 5;9)
(21) the bear got licked by the tiger
(22) he got bitten by the horse
(23) he got flied over by the tiger
(24) they got eaten by the horse
(25) he got rode on by Mickey Mouse
(26) he got riden on by Mickey Mouse
(27) Mickey got chasen oops... Pluto got chasen by Mickey Mouse and Minnie Mouse got chasen
(28) (the ball) got taken by the lady
(29) the ball got tooken by the lady
(30) it got droven by the man
(31) it got choppen down by me
(32) it got throwen by the lady
(33) it got aten by the cat

The frequency of the morphological forms used can be seen in Table 8. All tokens of the passive were included in the calculations, including repeated attempts. The AM children used more forms with *en*, fewer forms with *ed*, fewer bare stems, and more past tense stems than the SLI children.

As in Experiment 1, children were classified as *ed*-users, *en*-users or mixed. Three children in the AM group were classified as *ed*-users, six as mixed and four as *en*-users. While all of the children produced some correct regular *ed* forms, three *en*-

users and three 'mixed' children overgeneralized *en* to the regular *-ed* verbs. Examples of the various patterns can be seen in Table 9.

Summary of Results

Experiment 2 indicated that SLI children show a deficit in morphological analysis skills when compared to their age-matched peers; the two groups' performance differed significantly on almost all tasks.

Discussion

The results of the two studies indicate that SLI children show no deficit in the acquisition of syntax, but do have difficulty with the acquisition of inflectional morphology. Furthermore, SLI children show a deficit in morphological analysis skills when compared to their age-matched peers, but not when compared to their language-matched peers. The results of each task will be discussed below, followed by a general discussion of the implications of the two studies.

Passive Elicitation

Consistent with children reported by Crain and Fodor (1993) and Crain, Thornton and Murasugi (1987), the majority of the children were able to produce syntactically correct passives in the elicitation task. Syntactic errors, such as the lack of movement, shown in (16) below, were not found in the data.

(16) * got licked the bear

The productions can be considered true verbal passives. The presence of a prepositional phrase confirmed this for many children. For those children who did not produce full passives, the elicitation procedure provided the appropriate context for their interpretation as verbal passives. Given the elicitation question "what happened to X?," an adjectival response was not an appropriate response. The children demonstrated their knowledge of this fact by responding in the active voice if a passive was not elicited. They did not provide an alternative description of the patient, as might be expected in place of an adjectival passive. They were clearly attending to action rather than to description.

The vast majority of the passives elicited contained the verb *got* rather than *be*, consistent with the findings reported by Crain et al. (1987) and Crain and Fodor (1993). Nevertheless, some of the children (normal and SLI) did produce be-passives. The predominance of get- over be-passives might be because get-passives could be considered somewhat easier, due to the simpler morphological paradigm of *get* compared to *be*.[4] Alternatively, the fact that *get* in passives can be considered a main verb (Haegeman, 1985; Hoshi, 1991; Fox & Grodzinsky, 1992, Lasnik & Fiengo, 1974) might make them easier for children, since auxiliary verbs are known to be a source of difficulty (Brown, 1973; Johnston & Schery, 1976).

In spite of the large number of passives elicited in this study, some children produced only active sentences. The failure to elicit passives from these children cannot be attributed to age. The children who failed to produce passives were scattered throughout the age range of the language-matched group and included the two oldest children in the sample. The lack of passive production cannot be interpreted to mean that the children could not produce passives, only that they did not. As outlined in the methods section, the elicitation procedure sometimes required numerous attempts before meeting with success. Each child was given three separate opportunities, on different days, to produce the passive. In many cases, all three sessions were necessary. Note that these sessions did not teach the child the passive. The experimenter never used the passive structure during the task. The repeated sessions merely offered more opportunities for the passive to be elicited. Perhaps the remaining children would have produced passives, if given additional opportunities.

The ability of the SLI children to produce syntactically correct passives is consistent with the earlier findings that SLI children are capable of producing complex syntactic structures (Smith, 1992) and with Clahsen's (1989) claim that German SLI children do not suffer from a syntactic deficit. In spite of obvious difficulties in the acquisition of language, these children were able to produce passive syntax as well as their peers. This finding clearly supports the proposal that SLI children have an intact UG.

The children's proficiency with passive syntax is in sharp contrast with their lack of proficiency with the idiosyncratic linguistic structures stored in the lexicon, specifically prepositions and passive morphology. Prepositional errors were not uncommon and very few of the verbs elicited in the passive structure contained the correct inflection. In several cases, no overt passive morphology was present although the rest of the structure was grammatically correct. Since the affixation of the passive morphology is said to create the conditions which require the syntactic movement to take place (i.e., absorption of case and theta-roles), and since movement appears to have taken place, the inflection must have been added syntactically (perhaps as a null morpheme), but not realised morphologically. Such structures show clearly the distinction between (at least these) syntactic and morphological operations. As such, they offer support for a notion of syntactic inflection, realised in a separate part of the grammar from overt morphology.

In spite of the large number of errors, the morphology produced by all three groups of children showed impressive variety and creativity. Confusion with the two passive inflections, as well as the correct stem forms, was evident in the variety of irregular verb forms produced by all three groups of children. It is important to note that, in spite of the large number of errors, a form of *passive* morphology was always used. None of the children affixed a different inflection, such as /s/. This indicates that the children know which inflections affix to which categories. Further, it indicates knowledge of the special role of passive morphology and its effects on the syntax (i.e., manipulation of case and theta roles). The children do not attribute such characteristics to all inflections. This is consistent with the hypothesis of delay rather than deviance in SLI grammar.

A developmental trend in the use of *-ed* and *-en* in the three groups of children can be inferred from the cross-sectional data. The children showed a trend from 1) the use of *-ed* with overgeneralization to *-en* verbs, 2) the introduction of *-en* resulting in a variety of forms with either or both endings, plus occasional overgeneralization of *-en* to the *-ed* verbs, 3) appropriate use of *-en* (thus, eliminating the overgeneralization of *-ed*), with continued overgeneralization of *-en* to *-ed* verbs. Thus, there appears to be cross-sectional evidence for an overgeneralization paradigm, with both passive morphemes being over-used at times.

The frequency of overgeneralization in this study differs from the findings of Marcus, Pinker, Ullman, Hollander, Rosen, and Xu (1992) that overgeneralization occurs rarely in the speech of young children. Marcus et al. based their findings on the analysis of spontaneous speech transcripts collected as longitudinal studies of individual children over several years of their language development. The difference in the two studies' results might be attributed to the design differences. As Marcus et al. point out, it is possible that the use of an elicited production technique primes the children to produce overgeneralization. This may have happened in the current study due to the use in the elicitation protocols of the regular past tense (in the passive elicitation) or the bare stem (in the language screening). Nevertheless, priming cannot account entirely for the data, especially for the overgeneralization of *-en*. Neither the use of *-ed* nor a bare stem would be likely to prime a child to produce *chasen* instead of *chased*, a common error.

The cross-sectional, group design of this study may contribute to the differences. This study examined the use of the same 20 verbs (ten in the past tense screening task and ten in the passive task) in 49 children, 33 of whom were at the same level of morphological development. Thus, the sampling error encountered by Marcus et al. in their attempts to examine productions of the same verbs at one period of time was diminished. This study provided more data of a comparable type than the longitudinal transcripts studied by Marcus et al.

The pattern of overgeneralization accompanied by the inconsistency and variability in the verb forms used by the children illustrates the many different rules the children can hypothesise and the very active, almost experimental, approach these children are taking to the acquisition of passive morphology. This contrasts sharply with the lack of error in their production of the passive syntax. Thus, the answer to question (1), "Do SLI children show greater ability with structures based on the principles of UG than with idiosyncratic structures specific to a particular language?," is clearly "yes." At least with the structures studied here, the principles of UG appear to be intact in SLI children.

Morphological Analysis Tasks

On almost every morphological analysis task, the SLI children performed significantly worse than their age-matched peers and exactly the same as their language-matched peers. Each task will be discussed below.

Sentence completion tasks

As outlined above, the SLI children performed the same as their LM peers on this task. While all of the children could correctly complete some of the sentences, the overall performance was somewhat lower than one might expect from a sentence completion task. This difficulty can be accounted for by the morphological analysis demands of the task. Most sentence completion tasks provide an uninflected form and require the child to complete the sentence with an inflected form. For example, in the following item from the

Berry-Talbott Language Test (Berry & Talbot, 1966), the nonsense word *ling* is introduced uninflected, is then inflected and then the child is required to inflect it in a different way. "This is a tass who knows how to *ling*. He is *linging*. He did the same thing yesterday. What did he do yesterday? Yesterday he ________." The task used in this study provided only an inflected form of the non-word. The above item would have been presented in the following form: "This guy is *linging*. Yesterday he _____." Thus, the child in this study had to note that the verb was inflected with *ing*, the stem was *ling* and that the correct inflection to add was /d/.

The morphological analysis demands of the task are reflected in the type of errors the children made. The most common error was the repetition of the verb with the inflection used in the stimulus. The child making this error had adequate verbal memory skills to remember the exact form of the stimulus item but was unable to perform the morphological analysis necessary to separate the inflection from the stem. The lack of group differences on this error type indicated equal verbal memory skills on this task, in all three groups. Omission of the inflection also occurred, indicating adequate attention to the stem, but an inability to determine and add the correct inflection. The use of an incorrect inflection, another common error, reflected attention to the stem and the knowledge that an inflection was necessary, but an inability to analyse the grammatical context well enough to determine the appropriate inflection to be added.

Comprehension of inflected non-words

The comprehension task showed no differences between any of the groups on a straightforward comparison of mean scores. However, while only one child could be considered successful at the task in the LM group, four SLI and eight AM children could do the task. It appears that the ability to do this task begins to develop in normal children as they approach six years of age. This corresponds to the age at which children develop the ability to do many metalinguistic tasks (Liberman, Shankweiler, Fischer, & Carter, 1974). Most of the SLI children, on the other hand, were unable to do the task at age six. Most of those who were successful were in grade one and therefore had had some reading and writing instruction. While the direction of causation is not clear, reading and writing skills are correlated with morphological awareness (Carlisle, 1988; Rubin, 1988) and may have fostered awareness in these children. The comment made by one child (MD, SLI, 6;8) after a plural stimulus item, "it has an 's' at the end" is consistent with this hypothesis.

The difficulty that this task posed for these children deserves comment. One might have expected this to be a rather straightforward test of productivity of the plural inflection. Certainly, children both comprehend and use the plural marker consistently early in the acquisition sequence (Brown, 1973; Miller & Ervin, 1964). It is possible that the children in this study did not fully understand what was required of them. However, six training trials were provided, with feedback, in order to teach them the task. Further, the instructions emphasized that number was important, pointing out that one section contained one item, while the other contained two. Another possibility is that the children did not understand the question the way it was asked. However, a pilot study varied the instructions in many ways, with no effect. Finally, the training trials all contained the /əz/ allomorph because it was believed to be the most salient, while the test items included /s/ and /z/ as well. It is possible that the children did not generalise the training with /əz/ to the test items with /s/ and /z/. However, the children who were able to do the training trials correctly, (those who received a score of 6 out of 6) were also able to do the rest of the task, indicating that the training did generalise.

Judgment Task

The judgment task was very successful in eliciting judgments, identifications and repairs of morphological errors from very young children. It appears that normal 3-year-olds are quite capable of metalinguistic reflection of grammatical form.

The use of repetitive trials with feedback significantly increased performance in all three groups, particularly with respect to repairs in the SLI group. This increase in performance cannot be attributed solely to the children learning the procedure of the task. If the improvement were attributable to procedure learning, one would expect to see better performance on later items than on earlier items. This, however, was not the case. The improvement cannot be solely attributed to the scoring system either. A child was given credit for a response on trial 2 and 3 if she was correct on trial 1, possibly artificially inflating the scores on later trials. Nevertheless, an increase in scores across trials would only occur if children who were incorrect on earlier trials were correct on later trials. Thus, it appears that the children improved in their ability to detect and repair errors. This improvement over trials indicates that it is possible to teach children to do metalinguistic tasks. In this study, even minimal training improved performance, within the constraints of the child's language level. This is consistent with the findings of more extensive training studies of phonological awareness (Ball & Blachman, 1988; Bradley & Bryant, 1983, 1985; Lundberg, Frost, & Peterson, 1988; Warrick, Rubin, & Rowe-Walsh, 1993). This improvement with training has clinical and academic implications, given the relationship between morphological awareness and good reading and writing skills (Carlisle, 1988; Rubin, 1988). If normal and SLI children can be taught these skills, perhaps their reading and writing would benefit.

It is noteworthy that the SLI children benefited as much from the training as the LM children, and at times more. One might have expected the SLI children to be less receptive to teaching of language skills. Nevertheless, it appears that they benefit from training, at least within the limits of their expressive language abilities. Here again, the added reading and writing instruction the SLI children have received might have played a role.

The significant difference between performance on judgment, identification and repair is consistent with the findings of previous research (Smith-Lock & Rubin, 1993; Warrick et al, 1993; Warrick & Rubin, 1992). It appears that difficulty increases from judgments, to repairs to identifications, particularly by trial 3. All three groups demonstrated this pattern, although the AM group reached a ceiling on the second and third trials. One might expect the judgment task to be the easiest for several reasons. First, it required only a yes or no response. Chance alone would allow the correct answer 50% of the time. Second, it required minimum metalinguistic reflection. The child had only to determine if the sentence matched what she would say (i.e., did it match the output of her grammar?)

Repairing the error was somewhat more difficult. One might think that the repair of an error would simply involve the child spontaneously generating the correct sentence. In this task, that would mean commenting on the situation reflected in the toys still in front of the child. This may contribute somewhat to the relatively easy nature of this task. However, many children did make errors on the repairs. The difference in performance on judgments and repairs indicates that children sometimes correctly rejected the sentence, but were unable to repair the error. While some of these incorrect responses were no responses ("I don't know"), many of the errors involved a repetition of what the puppet had said, rather than a correction (again demonstrating good verbal memory abilities). This type of response reflects the metalinguistic demands of the task and the inability of the child to manipulate consciously what she has heard to produce a grammatical alternative.

The identification of the error was clearly the most difficult for the children. While judgment required a global comparison of the stimulus sentence to the child's grammatical output, and repair required the generation of such output, identification required the child to analyse each component of the sentence, identify which grammatical requirements were not met and then say the erroneous word/phrase aloud. As such, it was the most removed from an automatic speech task and involved a high amount of metalinguistic skill.

More difficult than any of the levels in the judgment task was the generation of morphological errors. The lack of a significant difference in the performance of the three groups reflects the low scores obtained by all. This task clearly demanded the most of the children. In order to be successful, they had to understand how the puppet had been grammatically manipulating the stimuli, and be able to identify and omit the inflectional morpheme themselves. The types of errors the children made shed light on their perception of the task. Many made no change at all, clearly lacking enough insight to even attempt a response. Semantic errors, of the sort *one eye* instead of *two eye*, demonstrated that the children were aware of the semantic implications of the change, but did not associate them solely with the inflectional morpheme. Phonological changes, most frequently substitutions of the initial phoneme, indicated that the children understood that a single segment was being manipulated. However, they did not comprehend the morphological significance of the segment or were unable to identify the inflectional morpheme in the stimulus.

GENERAL DISCUSSION

The answer to question (1), "Do SLI children show greater ability with structures based on the principles of UG than with idiosyncratic structures specific to a particular language?" is clearly "yes." The SLI children showed proficiency with the principles of theta-theory, case theory, and A-chains. In sharp contrast with this, they made many errors with passive morphology, producing few correct passive participles and overgeneralising *-ed* and *-en* . It can be concluded that, at least with respect to these grammatical principles, SLI children do not show a grammatical deficit. Furthermore, their morphological errors are not qualitatively different from those made by normal children. Thus, their language cannot be considered deviant in any way.

With respect to question (2), "Can SLI children's difficulty with inflectional morphology be attributed to a deficit in morphological analysis skills?"; the answer appears to be "yes" and "no". SLI children do demonstrate a deficit in morphological analysis skills with respect to their age-matched peers. However, they do not demonstrate a deficit in morphological analysis skills with respect to their peers matched on the basis of linguistic performance.

There are many possible reasons that the performance of the LM and SLI groups did not differ significantly. First, it is possible that the children in the SLI group were not truly SLI and, therefore, not representative of the population the study meant to tap. While all the SLI children had been referred by certified speech-language pathologists, due to the large number of sources from which the children were drawn, the same formal measures were not available for each child. However, data from this study do confirm their SLI status. The range of ages in the SLI and LM groups did not overlap and the mean ages of the groups differed by 2 years, 2 months. Thus, the SLI children showed a two-year delay in language level. Further, Experiment 2 confirmed that the SLI children performed significantly worse than children of the same age. These facts support the language-impaired diagnosis. That their deficit is specific to language is supported by their normal performance on a test of non-verbal intelligence. Thus, consistent with their independent diagnosis, the SLI group showed approximately a two-year delay in linguistic development, coupled with normal non-verbal skills, meeting the criteria for SLI.

Another possible explanation of the lack of differences is that the children were not adequately matched for language. If the SLI children were actually at a higher language level than the normal children, but still suffered a deficit in morphological analysis, they might have performed the same as the LM group, but below what would be expected for their language level. Nevertheless, there is little reason to believe that the language-matching was inadequate. To the contrary, the procedure appears to have been extremely successful. On the basis of ten verbs, two groups of children of differing ages and educational levels, drawn from six separate sources, were so well matched that they performed the same on all of the tasks.

The lack of differences between language-matched groups has been found by other researchers. Rubin, Kantor and Macnab (1991) found that SLI children aged 8;2 to 12;4 performed the same on grammatical analysis tasks as younger children matched on the basis of formal language testing. The lack of differences is also consistent with the findings of a study which matched children on the basis of written language (reading) level (Bryant & Impey, 1986). Bryant and Impey (1986) found that when dyslexic children were compared to normal children of the same reading level, the apparently deviant characteristics of the dyslexics disappeared. Normal readers were found to make the same errors with the same frequency as dyslexic children of a comparable reading level.

Another possible explanation of the data is that the tasks used in this study do not adequately assess morphological analysis skills. Thus, SLI children might suffer a deficit in morphological analysis which was not tapped in this study. The children's performance on these tasks argues against this, however. As discussed earlier, the sentence completion tasks differed from spontaneous speech in their analytic demands and reduced automaticity. This was supported by the preponderance of repetition errors in the data. The comprehension task clearly tapped morphological analysis skills, as reflected by its difficulty, by the spontaneous comments of the children and by the age (6 years) at which the children were able to do the task. The judgment task, a commonly used metalinguistic task, asked children to comment overtly on language, and the self-generated errors asked them to manipulate language in play. It appears that the tasks were successful in tapping morphological analysis skills. Thus, there must be another explanation of the results.

The level of language development of the children in the study may have contributed to the results. The children were specifically selected so that they had acquired the inflectional morphology system. Thus, only those children who had enough analysis skills to learn the morphological system and use it consistently were included. Perhaps the use of "first use" as the criterion for acquisition (as suggested independently by Stromswold, 1990, for normal children) would have produced different results. If children who demonstrated grammatical competence with the past tense (by the "first use" criterion) but not consistent grammatical performance, were studied, a difference might be found. In such a study, SLI children might show a deficit in morphological analysis skills at the time of "first use" of an inflection when compared to LM peers. A difference in analysis skills at that level of language development would lead to a protracted time to reach adequate performance levels on the part of the SLI children. Thus, the normal children would be expected to achieve consistent performance earlier than the SLI children. The attainment of adequate performance skills would coincide with the development of the necessary morphological analysis skills, leading to the results obtained in this study.

The equality of morphological analysis skills in children with the same oral language development indicates that the levels of linguistic analysis skills are closely associated with expressive language level, as defined by consistent performance. Rather than being a secondary skill which develops after primary language development, linguistic analysis appears to develop hand in hand with expressive language and is measurable in children as young as three years of age.

The role of linguistic analysis skills in primary language acquisition, as measured by the tasks in this study, must now be re-considered. Given the results, it is possible that linguistic analysis skills play no role in language development and, as such, language acquisition and linguistic analysis can be viewed as completely independent skills. However, the evidence does indicate that a close association between primary language skills and linguistic analysis skills exists. While it is possible that linguistic analysis skills as defined here play no role in the acquisition of grammatical competence, the possibility remains that these skills play a role in the attainment of consistent performance. Such a role, in a sense, reinforces the original view of linguistic analysis/awareness as a secondary skill which can be applied to the primary linguistic system (as outlined by Mattingly, 1972). However, in this view, the primary system could be considered the system involved in the acquisition of grammatical competence, while linguistic analysis skills come into play after the attainment of competence in order to aid in attainment of consistent performance.

The fact that SLI children develop linguistic analysis skills as they develop expressive language, just as normal children do, coupled with the passive data, where the SLI children made the same morphological errors as normal controls, paints a clear picture of language delay rather than deviance in SLI. Not only do SLI children appear to have the same language as younger children, but they seem to have the same secondary mechanisms, such as linguistic analysis ability.

The study's findings provide counter-evidence for Gopnik's (1991a, 1991b) and Gopnik and Crago's (1991) proposal that a grammatical deficit in the form of absent morpho-syntactic features can account for the language of SLI individuals. Contrary to her predictions, these SLI children showed evidence of the use of features through extensive overgeneralization of both the past tense and passive inflections. Furthermore, although there were no group differences on the task of comprehension of inflected non-words, four children in the SLI group were successful at the task, again implying the presence of features in their grammar.

The ability of the SLI children to produce passive syntax as well as the AM children is not consistent with Leonard (1989) and Leonard et al.'s (1988) "surface account", which predicted that the SLI children would have particular difficulty with passives. The clearest evidence against the surface account of passives is the mastery of the passive syntax in the absence of passive morphology. This occurred in 10% of the LM and SLI productions. However, in most cases, the children showed use of both passive syntax and morphology, albeit incorrect morphology, making the issue less clear. Nevertheless, the SLI children's success with passive syntax in contrast to its morphology suggests that, to the extent that perception of low-phonetic-substance morphemes is necessary for the acquisition of passive, SLI children show the same perception of the linguistic input as normal children. The surface account would still hold, however, if it turned out that passive could be acquired without adequate perception of low-phonetic substance morphemes. An interesting test case for the surface account would be the acquisition of nominal forms such as *the destruction of Rome by the Barbarians*. Such forms involve the application of passive within a noun phrase, triggered by the addition of the derivational, syllabic morpheme *-ion*. The surface account would predict that the acquisition of such nominal forms would be easier than the acquisition of verbal passives, given that the critical morpheme in the nominals is not a low phonetic substance morpheme.[5]

It must be noted that, due to the exclusionary criteria for the diagnosis of SLI, the SLI population might be a heterogeneous group. Such heterogeneity might account for the contradictory findings of the various studies of SLI. Nevertheless, the SLI children in this study did not appear to differ qualitatively from the SLI children referred for but not included in the study, at least to the extent of their screening performance. The lack of difference between the SLI and the normal controls in this study as opposed to others might be attributed to the use of a more accurate language matching procedure. However, the generalisation of findings based on a language-matched rather than a randomly selected sample, must be made with caution.

CONCLUSION

The results of these studies indicate that SLI children suffer from a selective delay in the acquisition of inflectional morphology. They demonstrated no difficulty in the acquisition of a complex syntactic structure, but made many errors with the complex morphology of the same structure. The principles of Universal Grammar examined in these studies (case theory, theta-theory and argument-chains) were intact in the SLI children. Linguistic delay rather than deviance was supported by the fact that the performance of the SLI children on elicited production and on tasks of morphological analysis could not be distinguished from that of younger normal children. It appears that linguistic analysis skills develop hand in hand with primary language skills, both in normal and SLI children. The finding that children with equal performance abilities have equal analysis skills is consistent with the proposal that linguistic analysis skills play a role in the attainment of consistent linguistic performance (through on-line monitoring of production), if not in the attainment of grammatical competence.

REFERENCES

Baker, M., Johnson, K., & Roberts, I. (1989). Passive arguments raised. *Linguistic Inquiry, 20,* 219-257.

Baldie, B. J. (1976). The acquisition of the passive voice. *Journal of Child Language, 3,* 331-348.

Ball, E., & Blachman, B. A. (1988). Phoneme segmentation training: Effect on reading readiness. *Annals of Dyslexia, 38,* 208-225.

Benton, A. (1964). Developmental aphasias and brain damage. *Cortex, 1,* 40-52.

Berry, M. F., & Talbot, R. (1966). *Berry-Talbot Language Tests: Comprehension of Grammar.* Rockford IL.

Borer, H., & Wexler, K. (1987). The maturation of syntax. In T. Roeper & E. Williams (Eds.), *Parameter setting.* Dordrecht: Kluwer.

Bowey, J. A. (1988). *Metalinguistic functioning in children.* Victoria, Australia: Deakin University Press.

Bradley L., & Bryant, P. (1983). Categorizing sounds and learning to read: A causal connection. *Nature, 3012,* 419-421.

Bryant, P., & Impey, L. (1986). The similarities between normal readers and developmental and acquired dyslexics. In P. Bertelson (Ed.), *The onset of literacy: Cognitive processes in reading acquisition.* Cambridge MA: The MIT Press.

Brown, R. (1973). *A first language: The early stages.* Cambridge, MA: Harvard University Press.

Carlisle, J. (1988). Knowledge of derivational morphology and spelling ability in 4th, 6th and 8th graders. *Applied Psycholinguistics, 9,* 247-266.

Chomsky, N. (1981). *Lectures on government and binding.* Dordrecht: Foris.

Clahsen, H. (1989). The grammatical characterization of developmental dysphasia. *Linguistics, 27,* 897-920.

Clark, E. (1978). Awareness of language: Some evidence from what children say and do. In A. Sinclair, R. J. Jarvella, & W. J. M. Levelt (Eds.), *The child's conception of language.* Berlin: Springer-Verlag.

Crain, S., & Fodor, J. D. (1993). Competence and performance in child language. In E. Dromi (Ed.), *Language and cognition: A developmental perspective.* Norwood, NJ: Ablex.

Crain, S., Thornton, R., & Murasugi, K. (1987). Capturing the evasive passive. paper presented to the 12th Annual Boston University Conference on Language Development

Eisenson, J. (1972). *Aphasia in children.* New York: Harper and Row.

Fox D., & Grodzinsky, J. (1992). A-chains in children's passive and get as an unaccusative (raising) verb. paper presented to the 17th Annual Boston University Conference on Language Development.

Gopnik, M. (1990a). Feature-blind grammar and dysphasia. *Nature, 344,* 715.

Gopnik, M. (1990b). Feature blindness: A case study. *Language Acquisition, 1,* 139-164.

Gopnik, M., & Crago, M. B. (1991). Familial aggregation of a developmental language disorder. *Cognition, 39,* 1-50.

Guilfoyle, E., Allen, S., & Moss, S. (1991). Specific language impairment and the maturation of functional categories. paper presented to the 16th Annual Boston University Conference on Language Development.

Haegeman, L. (1985). The get-passive and Burzio's generalisation. *Lingua, 66,* 53-77.

Horgan, D. (1978). The development of the full passive. *Journal of Child Language, 5,* 65-80.

Hoshi, H. (1991). The generalized projection principle and its implications for passive constructions. *Journal of Japanese Linguistics, 13,* 53-89.

Ingram, D. (1981). *Procedures for the phonological analysis of children's Language.* Baltimore: University Park Press.

Johnston, J. R., & Schery, T. K. (1976). The use of grammatical morphemes by children with communication disorders. In D. Morehead & A. Morehead (Eds.), *Normal and deficient child language.* Baltimore MD: University Park Press.

Kamhi, A. G., & Koenig, L. A. (1985). Metalinguistic awareness in normal and language-disabled children. *Language, Speech and Hearing Services in the Schools, 16,* 199-210.

Lasnik, H., & Fiengo, R. (1974). Complement object deletion. *Linguistic Inquiry, V,* 535-571.

Lee, L. L. (1966). Developmental sentence types: A method for comparing normal and deviant syntactic development. *Journal of Speech and Hearing Disorders, 31,* 311-330.

Leonard, L. (1989). Language learnability and specific language impairment in children. *Applied Psycholinguistics, 10,* 179-202.

Leonard, L. (1972). What is deviant language? *Journal of Speech and Hearing Disorders, 37,* 427-446.

Leonard, L., Bortolini, U., Caselli, M. C., McGregor, K. K., & Sabbadini, L. (1992). Morphological deficits in children with specific language impairment: The status of features in the underlying grammar. *Language Acquisition: A Journal of Developmental Linguistics, 2,* 151-179.

Leonard, L., Sabbadini, L., Volterra, V., & Leonard, J. (1988). Some influences on the grammar of English and Italian-speaking children with specific language impairment. *Applied Psycholinguistics, 9,* 39-57.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology, 18,* 201-212.

Lundberg, I, Frost J., & Peterson O. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly, 23,* 263-284.

Marcus, G., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., & Xu, F. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development, 57* (serial no. 228).

Marshall, J. C., & Morton, J. (1978). On the mechanics of EMMA. In A. Sinclair, R. J. Jarvella, & W. J. M. Levelt (Eds.), *The child's conception of language.* Berlin: Springer-Verlag.

Mattingly, I. G. (1972). Reading, the linguistic process and linguistic awareness. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye.* Cambridge: The MIT Press

McCauley, R. J., & Swisher, L. (1984). Use and misuse of norm-referenced tests in clinical assessment: A hypothetical case. *Journal of Speech and Hearing Disorders, 49,* 338-348.

Menyuk, P., & Looney, P. (1972). A problem of language disorder: Length versus structure. *Journal of Speech and Hearing Research, 15,* 264-279.

Miller, W. R., & Ervin, S. M. (1964). The development of grammar in child language. In U. Bellugi & R. Brown (Eds.), *The acquisition of language.* Monograph of the Society for Research and Child Development, 29, 9-33.

Pinker, S. (1984). *Language learnability and language development.* Cambridge MA: Harvard University Press.

Pinker, S., Lebeaux, D. S., & Frost, L. A. (1987). Productivity and constraints in the acquisition of the passive. *Cognition, 26,* 195-267.

Rees, N. (1973). Auditory processing factors in language disorders: The view from Procrustes' bed. *Journal of Speech and Hearing Disorders, 38,* 305-446.

Rice, M. L., & Oetting, J.B. (1991). Morphological deficits of SLI children: A matter of missing functional categories? Paper presented to the 16th Annual Boston University Conference on Language Development.

Rubin, H. (1988). Morphological knowledge and early writing ability. *Language and Speech, 31,* 337-355.

Rubin, H., Kantor, M., & MacNab, J. (1990). Grammatical awareness in the spoken and written language of language-disabled children. *Canadian Journal of Psychology, 44,* 483-500.

Scarborough, H. S., Rescorla, L., Tager-Flusberg, H., Fowler, A. E., & Sudhalter, V. (1991). The relation of utterance length to grammatical complexity in normal and language-disordered groups. *Applied Psycholinguistics, 12,* 23-45.

Shipley, K. G., Maddox, M. A., & Driver, J. E. (1991). Children's development of irregular past tense verb forms. *Language, Speech, and Hearing Services in Schools, 22,* 115-125.

Smith, K. M. (1992). The acquisition of long distance wh-questions in normal and specifically language-impaired children. Paper presented to the Linguistic Society of America Annual Meeting, Philadelphia.

Smith-Lock, K. M., & Rubin, H. (1993). Phonological and morphological analysis skills in young children. *Journal of Child Language, 20,* 437-454..

Steckol, K. F., & Leonard, L. B. (1979). The use of grammatical morphemes by normal and language-impaired children. *Journal of Communication Disorders, 12,* 291-301.

Stromswold, K. (1990). *Learnability and the acquisition of auxiliaries.* Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Warrick, N., & Rubin, H. (1992). Phonological awareness: Normally developing and language delayed children. *Journal of Speech Language Pathology and Audiology, 16, 1,* 7-16.

Warrick, N., Rubin, H., & Rowe-Walsh, S. (1993). Phoneme awareness in language delayed children: Comparative studies and intervention. *Annals of Dyslexia, 43,* 153-173.

Wechsler, D. (1989). *Wechsler Preschool and Primary Scale of Intelligence-Revised.* San Antonio, TX: The Psychological Corporation.

Wasow, T. (1977). Transformations and the lexicon. In P. W. Culicover, T. Wasow, & A. Akmajian (Eds.), *Formal syntax.* New York: Academic Press.

FOOTNOTES

[†] Also University of Connecticut

[1] Mean length of utterance is calculated by counting the number of morphemes in a spontaneous speech sample and dividing by the number of utterances, to determine the average number of morphemes per utterance. It is a rough indicator of linguistic development (Brown, 1973).

[2] This child used plural, possessive, present (third person singular) and past tense inconsistently in spontaneous speech. In spite of his inconsistent use of the regular past tense, he overgeneralised on two out of the ten screening verbs, indicating that "first use" might be a more appropriate measure of acquisition than "consistent use".

[3] One child was unable for testing on the passive task.

[4] I am grateful to Ignatius Mattingly for this suggestion.

[5] I am grateful to Mamoru Saito for this suggestion.

APPENDIX A

Screening Stimuli

irregular verb	age acquired*
went	3;6-3;11
saw	4;0-4;5
broke	5;0-5;5
took	5;0-5;5
fell	5;0-5;5
found	5;0-5;5
came	5;6-5;11
threw	5;6-5;11
made	5;6-5;11
sat	5;6-5;11

*Age at which 80% of children have acquired irregular past tense (Shipley, Maddox, & Driver, 1991)

APPENDIX B

Characteristics of SLI Children Included in the Study

The screening score is the number of incorrect irregular verb forms the child used in the screening task. Repeated attempts that resulted in use of both correct and incorrect forms were omitted from this score. Repeated attempts have, however, been included in the calculation of the number of overgeneralized and uninflected stems. Thus, the total of the overgeneralizations plus uninflected stems might be higher than the screening score. Spontaneous productions of irregular verbs not in the screening protocol were not included.

child	age	screening score	overgeneralization	uninflected stem
AP	7;2	6	2	5
IP	6;0	6	5	1
CJ	5;4	7	5	2
BE	6;5	8	4	4
MK	7;3	6	4	2
JH	5;7	6	2	5
JG	5;9	6	8	1
RZ	5;9	6	3	2
KG	6;0	6	5	1
JC	5;4	6	5	1
TK	6;9	6	7	0
MQ	5;10	7	5	2
HB	5;4	7	6	1
AG	6;11	6	7	1
MD	6;8	5	3	2
JM	6;7	5	2	3
SC	5;11	8	2	6

Characteristics of LM Children

child	age	screening score	overgeneralization	uninflected stem
MW	4;2	5	8	0
PR	3;11	6	3	3
JP	4;1	9	4	5
SW	4;1	7	5	5
KM	3;11	6	4	2
DC	4;5	5	2	3
AT	3;11	6	6	1
DM	4;1	6	3	3
AM	3;11	5	5	1
AF	4;2	6	4	2
KH	4;2	5	3	2
SW	4;0	5	1	4
CG	3;11	7	5	3
BM	3;4	8	5	3
AB	4;1	8	5	3
LS	4;3	6	5	2

Haskins Laboratories Status Report on Speech Research
1993, SR-114, 139-158

Hemispheric Asymmetries in Adults' Perception of Infant Emotional Expressions[*]

Catherine T. Best,[†] Jane S. Womer,[‡] and Heidi Fréya Queen[‡]

Accounts of emotion lateralization propose either overall right hemisphere (RH) advantage, or differential RH vs. LH involvement depending on the negative-positive valence of emotions. Perceptual studies generally show RH specialization. Yet viewer emotional responses may enhance valence effects. Because infant faces elicit heightened emotion in viewers, we assessed perceptual asymmetries with chimeric infant faces. First, we determined that chimeras must be paired with their counterparts, not their mirror-images, to tap viewers' sensitivity to adult facial asymmetries. Next, we found an RH perceptual bias for infant cries, but bihemispheric sensitivity to asymmetries in infant smiles. This effect was not due to LH featural vs. RH holistic processing, and held for additional, intensity-matched, spontaneous expressions. Specialized RH sensitivity to infant cries may reflect an evolutionary advantage for rapid response to infant distress.

Effects of Emotional Valence and Hemiface Differences on Adult Perceptual Asymmetries for Infant Facial Expressions

Findings with both unilateral brain-damaged patients and normal adults have led to general consensus that the human cerebral hemispheres are differentially involved in emotional, as well as cognitive, processes. However, the exact pattern of hemispheric involvement in emotions remains controversial. According to the most widely-held view, the right hemisphere (RH) dominates overall in perception and expression of emotion, across both negative and positive valence (e.g., Campbell, 1978; Chaurasia & Goswami, 1975; Gainotti, 1972, 1988; Hirschman & Safer, 1982; Ladavas, Umilta & Ricci-Bitti, 1980; Ley & Bryden, 1979, 1981; Safer, 1981; Strauss & Moscovitch, 1981). For convenience, we will refer to that view as the ***RH hypothesis*.** The major counter-proposal has been that the RH predominates in negative emotions, the left (LH) in positive, a view we will call the ***valence hypothesis*** (e.g., Ahern & Schwartz, 1979; Dimond & Farrington, 1977; Natale, Gur & Gur, 1983; Reuter-Lorenz & Davidson, 1981; Reuter-Lorenz, Givis & Moscovitch, 1983; Rossi & Rosadini, 1967; Sackeim et al., 1982; Silberman & Weingartner, 1986; Terzian, 1964). Several variations on the valence hypothesis have also been offered. Some evidence suggests that while negative emotions show differential RH involvement, there may be less hemispheric asymmetry for positive emotions (e.g., Dimond, Farrington & Johnson, 1976; Ehrlichman, 1988; Sackeim & Gur, 1978, 1980); we will call this the ***negative-valence hypothesis***. Another possibility is that differential

---

This work was supported in part by NIH grants DC-00403 and DC-00045 to the first author, and by NIH grant HD-01994 to Haskins Laboratories. Thanks are due to the following people for their help in completing the paper and the research described here: Christine Blackwell, Shama Chaiken, Linda Creem, Angelina Diaz, Glenn Feitelson, Sari Kalin, Laura Klatt, Dara Lee, Roxanne Shelton, Alicia Sisk, Leslie Turner and Jennifer K. Wilson for their help with stimulus preparation and data collection and scoring; to Nathan Brody, Michael Studdert-Kennedy, James Cutting and two reviewers for helpful comments on earlier ms drafts; to Katherine Hildebrandt (now Karraker) for making her infant photographs and emotional rating data available; to Jerre Levy for lending her free-field chimeric test of smiling adults; and especially to the parents and staff of the Woodbridge and Yeladim daycare centers for their cooperation and help in getting the infant photographs for Experiment 5.

hemispheric involvement in emotions may depend on the motivational qualities of approach versus avoidance, rather than valence *per se* (e.g., Kinsbourne, 1978). According to Davidson and colleagues, the hemispheric approach-avoidance distinction pertains only to the subject's internal feeling-state and expressions (both mediated by frontal lobes), but not to perception of emotions (parietal lobes), which show an overall RH superiority (Davidson, 1984, 1992; Davidson & Fox, 1982; Davidson, Schwartz, Saron, Bennett, & Goleman, 1979; Fox & Davidson, 1986, 1987, 1988). We will call the latter proposal the ***motivational hypothesis***.

This report focuses on normal adults' perceptual asymmetries for infant facial expressions. Findings on perception of *adult* facial expressions by neurologically-intact subjects generally favor the RH hypothesis (e.g., Brody, Goodman, Halm, Krinzman & Sebrechts, 1987; Bryden, 1982; Bryden & Ley, 1983; Campbell, 1978; Carlson & Harris, 1985; Gage & Safer, 1985; Heller & Levy, 1981; Hirschman & Safer, 1982; Ley & Bryden, 1979, 1981; Moscovitch, 1983; Safer, 1981; Segalowitz, 1985; Strauss & Moscovitch, 1981). They have typically found a left visual field (LVF) advantage (RH superiority) for both positive and negative expressions.

A few perceptual studies have supported the other hypotheses. Favoring the valence hypothesis, adults rate tachistoscopically-presented facial expressions more negatively in the LVF-RH, more positively in the right visual field (RVF-LH), although the RH is better overall at differentiating emotions (Natale et al., 1983). Similarly, subjects detect which visual field contains a negative expression (vs. a contralateral neutral expression) more rapidly in the LVF-RH, but detect positive expressions more rapidly in the RVF-LH (Reuter-Lorenz & Davidson, 1981; Reuter-Lorenz et al., 1983). Supporting the motivational hypothesis, both adults (Davidson et al., 1979) and infants (Davidson & Fox, 1982) show greater EEG activation in frontal RH while viewing emotionally negative films, but greater LH frontal activation during positive films; parietal activation is greater in RH at both ages for both film types. Consistent with the negative-valence hypothesis, when emotionally negative films (Dimond et al., 1976) or odorants (Ehrlichman, 1988) are lateralized to a single hemisphere, subjects rate RH stimuli as more intensely negative, but fail to show asymmetries for rating positive stimuli.

Why the inconsistencies? One possibility is that studies favoring the RH hypothesis have often, though not always, assessed recognition or discrimination of facial expressions, whereas studies showing valence effects have called for judgments about stimulus emotionality. Recognition and discrimination can be carried out by so-called "cold" cognitive abilities, but emotionality judgments may encourage the viewer to tap into emotional processes. Perceptual asymmetries may be enhanced by the viewer's emotional response to the stimuli (e.g., Safer, 1981), perhaps especially to their valence properties (see Davidson, 1984; Ehrlichman, 1988). Emotional response may, in turn, be influenced by whether the expressions are spontaneous or posed. Spontaneous emotional expression is disrupted by temporal or extrapyramidal damage, posed expression by frontal or pyramidal damage (e.g., Monrad-Krohn, 1924; Remillard, Anderman, Rhi-Sausi & Robbins, 1977; Rinn, 1984). Spontaneous vs. posed expressions likely carry information about the emitter's emotional state. Perceivers should be more likely to respond emotionally to genuine than simulated expressions. Notably, most perceptual asymmetry studies have used posed rather than spontaneous stimuli.

Therefore, we conducted a series of studies involving emotionality judgments about stimuli that are highly likely to elicit emotional responses: smiling and crying infants. Infants' expressions are more spontaneous than adults', which are influenced by social conditioning and cultural display rules (Buck, 1986; Ekman, 1972). Those factors have little or no influence on young infants, who are thought not to simulate or mask emotional expressions until the second year (e.g., Campos, Barrett, Lamb, Goldsmith & Stenberg, 1983; Oster & Ekman, 1978; Rothbart & Posner, 1985; Sroufe, 1979; but see Fox & Davidson, 1988, and our Experiment 5). Moreover, adult expressions are said to often show complex emotion mixtures, making them more difficult to "read" than infant expressions, which are thought to display simple, basic emotions (Campos et al., 1983; Izard, 1979; Izard, Huebner, Risser, McGinnes & Dougherty, 1980; but see Oster, Hegley & Nagel, 1992). Most importantly, ethological research indicates that infant faces elicit stronger emotional responses in viewers than do those of (unknown) adults (e.g., Bowlby, 1969; Eibl-Eibesfeldt, 1975; Lorenz, 1935, 1981; Lorenz & Leyhausen, 1973). Adults' emotional responses to infant expressions are part of a mutually adapted behavior system that shapes communicative interactions, and that presumably evolved to promote nurturance and survival of the relatively helpless human infant. These responses are particularly strong in infants' caregivers, but are present in all humans.

But what role might perceptual asymmetries play in face-to-face interactions between adult and infant? Infant crying and smiling are of particular interest here. Both promote physical proximity between infant and caregiver, though for different reasons (e.g., Bowlby, 1969; Campos et al. 1983; Emde, Gaensbauer & Harmon, 1976). Infant smiling indicates a *positive* affective state and emotional *approach* toward the adult partner, and typically elicits corresponding *positive* feelings and *approach* from the adult. Infant crying, however, indicates *negative* feelings toward a noxious stimulus, and thus a *withdrawal tendency*. Infant crying typically evokes *negative* feelings of concern in adults, who usually want to *approach* in order to mitigate the infant's distress. This analysis leads to different predictions by the four hypotheses regarding cerebral organization for emotional processes. The RH hypothesis predicts an overall RH bias unaffected by valence. The valence hypothesis predicts a RH advantage for infant crying expressions, but a LH advantage for smiles. The negative-valence hypothesis predicts a RH advantage for cries only. The motivational hypothesis should predict a *LH* advantage for cries and smiles, both of which motivate approach responses in adults.

To investigate these possibilities, we tested perception of photographs of smiling and crying infants. A free-viewing procedure (Levy, Heller, Banich & Burton, 1983a) was deemed best suited to the ecological condition of interest—adults' perception of infants in natural face-to-face situations—and to future studies with infant and child viewers (see Levine & Levy, 1986). For each page of the test booklet, subjects chose which of two half-neutral/half-emotional chimeras of a given infant appeared happier (or sadder); the emotional expression was on the left in one chimera, on the right in the other, with top vs. bottom position on the page counterbalanced across items. Because binary forced-choice data avoid floor and ceiling effects, performance level corrections such as the Phi coefficient (Kuhn, 1973) or λ (Bryden & Sprott, 1981) are neither necessary nor applicable; unbiased asymmetry scores on this task are obtained via simple laterality ratios (Levine & Levy, 1986).

Visual asymmetries have often been assessed via tachistoscopic lateralization of stimulus input to a single hemisphere, under the assumption that tasks involving lateralized input provide a more precise and controlled index of hemispheric processing differences than do free-field tasks. However, theory and recent findings question this assumption. Over two decades ago, Kinsbourne (e.g., 1970, 1978) argued that task requirements and subject expectancies increase the activation of the hemisphere that the subject employs preferentially for the perceptual or cognitive functions involved. This biases attention to the spatial hemifield contralateral to the more active hemisphere, which heightens sensitivity to, and perceived intensity of, stimuli in that hemifield and results in the lateral asymmetries observed in both free-field and lateralized-input tasks. Although there were some failures to replicate certain of Kinsbourne's specific results, recent findings with brain-damaged and intact adults support the claim that activational asymmetries cause perceptual biases in tachistoscopic tasks (e.g., De Renzi, Gentillini, Faglioni & Barberi, 1989; Reuter-Lorenz, Kinsbourne & Moscovitch, 1990). In fact, tasks in which input is restricted to one hemisphere are subject to individual differences in attentional biases or hemispheric arousal asymmetries (Levy, Wagner & Luh, 1990; Mondor & Bryden, 1992). Moreover, a number of free-field tasks find the expected left spatial field (LSF) bias in tests of RH functions (e.g., Levy et al., 1983a; Luh, Rueckert & Levy, 1991) and a right spatial field (RSF) bias for LH functions (Levy & Kueck, 1986), and do so reliably (Wirsen, Klinteberg, Levander & Schalling, 1990). Indeed, subjects' free-field perceptual biases are predicted by their asymmetries on tachistoscopic tasks (Burton & Levy, 1991; Kim, Levine & Kertesz, 1990; Hellige, Bloch & Taylor, 1988; Wirsen et al., 1990). The correlation reflects individual variations in characteristic arousal differences between the hemispheres (e.g., Levy, Heller, Banich & Burton, 1983b), which are corroborated by individual differences in EEG alpha asymmetry in the parietal and temporal regions (Green, Morris, Epstein, West & Engler, 1992), which include the cortical projection area of the posterior, visuo-spatial attention system (Posner & Peterson, 1990).

On the chimeric free-field task, a LSF-RH bias for negative infant expressions would be expected according to the RH, valence, and negative-valence hypotheses; however, the motivational hypothesis predicts a RSF-LH bias. The four theoretical models differ as to whether infant smiles should yield a LSF-RH bias (RH hypothesis), a RSF-LH bias (valence and motivational hypotheses), or no asymmetry (negative-valence hypothesis).

Moreover, the heightened perceptual sensitivity to infant expressions that is predicted by ethological theory suggests that viewers' perceptual asymmetries should also be influenced by hemiface differences in infant emotional expressiveness. Infants, like adults, show greater emotional intensity on one hemiface, a manifestation of hemispheric differences in expression of emotions. Unlike adults, however, who show a left hemiface expressive bias, infants show a right hemiface bias (Best & Queen, 1989; Rothbart, Taylor & Tucker, 1989). Therefore, we wanted our task to detect interactions between infant hemiface biases and adult perceptual asymmetries. In their original tachistoscopic study with chimeras of smiling adults, Heller & Levy (1981) found just such an interaction between emitters' hemiface biases and viewers' perceptual asymmetries. However, their free-field test with the same faces (Levy et al., 1983a) failed to find a hemiface effect on perception. While the tachistoscopic measure might be more sensitive than the free-field one, differences in the construction of the chimeric choice pairs in the two studies provide another potentially important methodological factor. Each chimera in the tachistoscopic study was paired with one generated from the other halves of the same photos. The free-field task instead paired each chimera with its mirror-reversed print. Thus the pairs in the tachistoscopic task retained information about emitter hemiface asymmetries, whereas those in the free-field task did not. To determine whether the free-field task *can* detect emitter hemiface effects on perceptual asymmetries, we first tested two versions of the Levy et al. free-field adult face task, which differed only in how the chimeric choices were paired.

As indicated earlier, the expressive asymmetries of the smiling emitters in the Levy et al. test booklet had been previously determined tachistoscopically in Heller and Levy (1981), via viewers' paired-comparison emotionality judgments of mixed-expression chimeras of each emitter. Because our interest was in perceived emotionality, perceptual evidence about the expressive asymmetries of the stimulus faces was deemed most appropriate for our purposes (as opposed to, e.g., taking some physical measurement of each hemiface, which may not necessarily map straightforwardly to perceived emotionality of the two hemifaces—see also footnote 1). Although not all of the emitters in the Levy et al. test booklet had shown a left hemiface bias in smiling, we used their full set of stimuli because we needed to replicate their findings for comparison against the results from free-field presentations of the pairings used in the Heller and Levy (1981) tachistoscopic study.

EXPERIMENT 1

Method

Subjects. The subjects were familial right-handers, who show stronger, more consistent cerebral asymmetries than non-right-handers, including emotion perception asymmetries (Chaurasia & Goswami, 1975; Heller & Levy, 1981). The handedness checklist assessed degree of hand preference on 10 unimanual activities, as well as writing hand of immediate family members. Right-handedness was defined as a "strong" to "moderate" right-hand preference for all items, with no switch during childhood, and both parents right-handed. Four subjects failed to meet these criteria. Subjects were university students with normal or corrected vision, who received $4.00 for participation. Forty-six subjects (23 male, 23 female) completed Test A (see *Procedure*), and 58 completed Test B (29 male, 29 female). All had participated in a related study of asymmetries in infants' facial expressions (Best & Queen, 1989).

Stimuli. We used the chimeras of half-smiling, half-neutral adult faces constructed by Heller and Levy (1981) from frontal photographs of nine young men, including both right- and left-handers, whose smiles had been elicited by the photographer's own smiling and joking. Given that the photographer was unfamiliar to the men, their smiles were most likely the socially conditioned sort rather than the truly spontaneous, genuine smiles that occur in interactions among good friends. All nine emitters displayed strong evidence of *orbicularis oculi* muscle activity, which causes cheek-raising, eye narrowing, and crinkling at the outer corners of the eyes (AU6-7 muscle involvement) and results in the appearance of "happy eyes." AU6-7 activity has been posited to occur only with smiles that are "felt", i.e., spontaneous and genuine expressions of heartfelt positive emotion; such "felt" smiles are claimed by some to show symmetry rather than asymmetry (Ekman & Friesen, 1982). Nevertheless, Heller and Levy (1981) found that all but one emitter had asymmetrical smiles; six were perceived to be more expressive on the left hemiface, two on the right.[1] The right-handed viewers in that study showed a LVF (RH) perceptual bias across this set of emitters.

The two normal orientation chimeras of each emitter had been made by joining the left half of the smiling photo with the right half of the neutral photo, and conversely, the right half of the smile with the left half of the neutral. Mirror-reversed chimeras were constructed from reverse prints of the photos.

Procedure. The Test A booklets were those developed by Levy et al. (1983). Each normal orientation chimera (9 emitters × 2 chimeras) was paired with its mirror-reversed counterpart. Each pair was presented one below the other on 8-1/2" × 11" pages, and appeared twice in the randomized 36-page test booklet, once with the normal orientation chimera at the top, once at the bottom. For the 36-page Test B booklets we re-paired the chimeras as in Heller and Levy (1981), such that each normal orientation chimera was presented with its normal orientation counterpart, and each mirror-reversed chimera was likewise presented with its mirror-reversed counterpart. Thus each choice pair in Test B retained evidence of hemiface differences between each emitter's two half-smiles, but those hemiface biases were missing from each Test A pair.

Subjects were run in small groups in a quiet windowless room, with Test A and Test B conducted as separate experiments. Each subject had a separate copy of the booklet. Their task was to write on their answer sheet which of the two items appeared happier for each page of their booklet. Test completion was self-paced, but subjects were told to follow their initial reaction rather than deliberating over their choices. They were told there were no correct answers, and that they should do the pages in order without comparing or changing answers.

Results

The data were converted to laterality ratios via the formula (R-L)/(R+L), in which R = percent of choices with the emotional expression on the right side of the chimera (i.e., RSF preference), L = percent choices with it on the left (LSF preference), and R+L = 100%. The laterality ratios thus range from -1.0 (LSF bias) to +1.0 (RSF bias). The Test A data were entered into 2 × 2 analysis of variance (ANOVA) for the factors of subject sex and emitter hemiface (i.e., whether the half-smile of the mirror-imaged chimera pairs was from the right or left hemiface of the emitter). Test B data were entered into a separate 2 × 2 ANOVA, for the factors of subject sex and chimera orientation (i.e., normal or mirror-reversed pairs). To determine whether specific laterality ratios showed significant asymmetry (i.e., deviation from 0), two-tailed *t*-tests were conducted, with alpha level correction for multiple *t*-tests set at $p < .0125$ for Test A and $p < .007$ for Test B.

In the Test A analysis, only the grand mean was significant, $F_{(1,\ 44)} = 15.62$, $p < .0003$, indicating a significant LSF-bias ($M_{\text{lat ratio}} = -.302$) in perceived intensity of the chimeric half-smiles. Neither sex nor hemiface nor their interaction was significant. The LSF effect was significant both when the emitter's left hemiface provided the smile ($M_{\text{lat ratio}} = -.294$), $t_{(45)} = -3.73$, $p < .0005$, and when the right hemiface did, ($M_{\text{lat ratio}} = -.309$), $t_{(45)} = -3.75$, $p < .0005$, and for both male viewers($M_{\text{lat ratio}} = -.23$), $t_{(22)} = -2.84$, $p < .01$, and female viewers ($M_{\text{lat ratio}} = -.374$), $t_{(22)} = -4.53$, $p < .0005$.

In the Test B analysis the grand mean effect, $F_{(1,\ 56)} = 19.19$, $p < .0001$, was also LSF-biased ($M_{\text{lat ratio}} = -.155$). Note, however, that it was only half the magnitude of that for Test A. Moreover, both the orientation effect, $F_{(1,\ 56)} = 248.94$, $p < .0001$, and the sex effect, $F_{(1,\ 56)} = 18.72$, $p < .0001$, were significant. The orientation effect indicated that the LSF bias occurred only for the *mirror-reversed* chimera pairs ($M_{\text{lat ratio}} = -.529$), $t_{(57)} = 12.51$, $p < .0001$; normal-oriented chimeras showed a significant *RSF* bias ($M_{\text{lat ratio}} = +.218$), $t_{(57)} = -4.25$, $p < .0001$. Males showed an overall LSF bias ($M_{\text{lat ratio}} = -.308$), $t_{(28)} = -8.19$, $p < .0001$, while females showed no overall asymmetry ($M_{\text{lat ratio}} = -.002$). While the sex × orientation effect was not significant, male viewers' striking LSF bias for mirror-reversed chimeras ($M_{\text{lat ratio}} = .-651$), $t_{(28)} = -19.06$, $p < .0001$, was met by a lack of significant asymmetry for normal-oriented chimeras ($M_{\text{lat ratio}} = +.034$), but females' LSF bias for mirror-reversed chimeras ($M_{\text{lat ratio}} = -.406$), $t_{(28)} = -9.23$, $p < .0001$ was opposed by an equally large RSF bias for normal-oriented chimeras ($M_{\text{lat ratio}} = +.402$), $t_{(28)} = 8.09$, $p < .0001$ (see Figure 1). That is, while both sexes were sensitive to emitter expressive asymmetries, this interacted with spatial hemifield asymmetries in male viewers, but instead it overpowered hemifield asymmetries in females. Emitter asymmetries enhanced or attenuated male viewers' perceptual bias, dependent on whether the more intense half-smile appeared in the more attentionally-biased hemifield, but stimulus asymmetry was apparently the sole determinant of female performance on Test B.

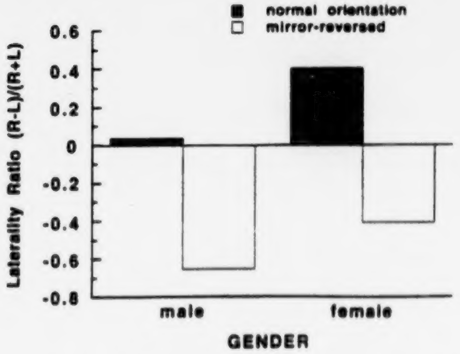


Figure 1. **Effect of chimera orientation on male vs. female viewers in Test B of Experiment 1 (smiling adult chimeras). In normal orientation chimera pairs, the left hemiface of the emitter's smile appears on the right; in mirror-reversed pairs, the left hemiface smile appears on the left. Negative laterality ratios indicate a left spatial hemifield bias, positive scores a right hemifield bias.**

DISCUSSION

The overall LSF bias found with the mirror-image pairings of Test A replicates the Levy et al. (1983) findings with the same booklet, and supports the RH hypothesis for perception of adult smiling faces. This result runs counter to the other three theoretical hypotheses, except possibly the motivational hypothesis, which assumes RH parietal involvement in simple perception of both negative and positive expressions.

However, the Test B results complicate this interpretation. When the chimera choices retain emitter hemiface differences, those *expressive* asymmetries significantly affect the viewers' perceptual asymmetries in this free-field task, just as in a tachistoscopic test (Heller & Levy, 1981). For normal orientation chimera pairs, the emitter's left-hemiface (LF) smile (the more expressive hemiface, on average) falls in the viewer's *less* sensitive RSF, but for mirror-reversed chimeras the more expressive LF smile falls in viewer's *more* sensitive LSF. Male viewers showed a trading relation between their basic LSF attentional asymmetry (tapped in Test A) and the emitters' LF expressive asymmetries. Cooperation between the two asymmetries in the case of mirror-reversed chimera pairs enhanced the magnitude of LSF bias in viewers' choices. But the two asymmetries were in conflict in the case of normal orientation pairs and thus cancelled each other's effects.

Female viewers, however, did not show this trading relation. Instead, their choices for normal vs. mirror-reversed chimera pairs showed equal-magnitude but directionally opposite biases, i.e., they depended exclusively on emitter expressive asymmetries. That their laterality ratios were not at the extremes of the possible range (-1.0 and +1.0) may reflect individual differences in the direction and degree of expressive asymmetry in the emitters, two of whom were reported to have *RF* expressive biases, another a complete lack of expressive asymmetry (Heller & Levy, 1981). The crucial point is that when expressive asymmetries were evident in the paired choices of Test B, for female viewers those expressive asymmetries apparently overpowered the effect of the basic attentional asymmetry that *was* evident in females on Test A. Thus, the two asymmetry factors interact in male judgments about stimulus emotionality, but stimulus asymmetry takes primacy over spatial hemifield biases in female judgments. Another possibility, though not mutually exclusive, is that the differential impact of emitter asymmetries may reflect sex differences in perceiving the smiles of young men.

In any event, the Test B approach is better suited to assessing how adult attentional asymmetries when viewing infant faces may interact with the infants' expressive asymmetries. Would infant expressions, like adults', elicit an overall LSF bias even for smiles, supporting the RH hypothesis? Or would the increased emotional response to infant faces result in a valence effect on attentional asymmetry? These questions were examined with emotional/neutral chimeras of smiling and crying infants, presented in a free-field task. To assess any interaction between infant expressive asymmetries and attentional asymmetries, we retained information about hemiface asymmetries in each chimeric choice pair as in Test B. Recall that our previous study of infant expressive asymmetries had found a right hemiface (RF) bias in infant cries and smiles (Best & Queen, 1989; also Rothbart et al., 1989), contrary to the LF bias found in adults' expressions. Thus, whereas in normal face-to-face interactions most adults' more expressive LH appears in the viewer's *less* sensitive RSF, most infants' more expressive RF appears in the viewer's more sensitive LSF. That is, the RH hypothesis predicts that for *normal* orientation chimeras the infant's expressive asymmetry and the viewer's attentional asymmetry will usually coincide, enhancing the LSF perceptual bias regardless of the emotional valence depicted. According to variants of the valence hypothesis, however, the pattern of asymmetries may differ for crying vs. smiling expressions, due to heightened emotional responses toward infants. Specifically, the valence and negative-valence hypotheses predict the same LSF pattern for cries as does the RH hypothesis. For smiles, however, the valence hypothesis predicts an *RSF* bias that is stronger for mirror-reversed than normal orientation, while the negative-valence hypothesis predicts an orientation-dependent shift in perceptual asymmetry concordant with the spatial position of the more expressive RF. Finally, the motivational hypothesis should predict an *RSF* bias for both the smiles and cries of infants, stronger for mirror-reversed than normal orientation chimeras; as argued earlier, both expressions should elicit an approach tendency from the viewer.

EXPERIMENT 2

Method

Subjects. The 46 subjects who took Test A in Experiment 1 also participated in this study.

Stimuli. The stimulus materials were generated from photographs of facial expressions by 10 normal, full-term 7- to 13-month-old infants, originally taken by a portrait photographer for a series of infant attractiveness studies (Hildebrandt & Fitzgerald, 1978, 1979, 1981). The same original photographs were used in Best and Queen (1989). In that study, viewers made paired-comparison judgments of mirror-image composites of each infant's left versus right hemiface. Their data indicated that the infants' showed more intense emotional expressions on the right hemiface than on the left; this was true for both smiling and crying expressions.

For the present study, each of those infants provided a neutral expression and either a clear-cut negative (crying) or a clear-cut positive (smiling) expression, according to ratings obtained in an independent study (Hildebrandt, 1983). Four infants had crying expressions; six had smiling expressions. Only two of the smiling infants displayed AU6-7 eye "crinkling" activity; these were the two youngest infants photographed. All photographs were full-frontal facial views.

Chimeras were constructed as in Experiment 1 (see Heller & Levy, 1981). Each print was cut exactly down facial midline, defined by a line extending through the point midway between the internal canthi of the eyes and the point in the center of the philtrum just above the upper lip. For each chimera, the hemifaces were aligned at the eyes and nose (mouths often could not be exactly aligned because of differing degrees of opening; see also Heller & Levy, 1981).

Each chimera was then centered behind an oval-shaped mattboard opening the size of the average photographed face, to screen out variations among infants in hair and facial outline. Copies were made with a high-quality Kodak photocopier, using a gray-scale photo correction template. Each infant was represented on four pages, as in Test B of Experiment 1. Thus, there were 40 pages of paired chimeras. The pages were ordered pseudorandomly, with no more than three consecutive smiling or crying infants, and no consecutive presentations of the same emitter. The question "Which infant looks happier?" (for smiling chimeras) or "Which infant looks sadder?" (for crying chimeras) was printed at top of each page.

Procedure. Testing was as in Experiment 1, except for the question valence difference.

Results

Laterality ratios were entered into a 2 × 2 × 2 analysis of variance (ANOVA) for the factors of emotion (cry, smile), orientation (normal, mirror-reversed), and sex. As before, *t*-tests were used to test significance of laterality ratios; the alpha adjustment was set to $p < .0065$.

There was a significant though modest LSF bias overall (M_{lat} ratio = -.13), $t_{(45)} = -4.78$, $p < .0001$. However, a significant emotion effect, $F_{(1,45)} = 10.09$, $p < .003$, indicated that valence influenced the asymmetry of the adults' judgments about the intensity of infant expressions. Specifically, the LSF bias was significant for crying infants (M_{lat} ratio = -.19), $t_{(45)} = -4.84$, $p < .0001$, but not for smiling infants (M_{lat} ratio = -.07). In addition, the orientation effect, $F_{(1,45)} = 366.68$, $p < .0001$, revealed that adult viewers' perceptual biases were sensitive to asymmetries in the infants' expressions themselves. Normal orientation chimeras, in which the infants' more expressive RH (Best & Queen, 1989) appeared in the LSF, yielded a significant LSF bias (M_{lat} ratio = -.57), $t_{(45)} = -17.84$, $p < .0001$, whereas mirror-reversed chimeras, in which the infant RH appeared in the RSF, yielded a smaller but significant RSF bias (M_{lat} ratio = +.31), $t_{(45)} = 8.02$, $p < .0001$. Finally, the significant emotion × orientation interaction, $F_{(1,45)} = 66.80$, $p < .0001$, found that laterality ratios for smiles reversed from a strong LSF bias for normal orientation chimeras (M_{lat} ratio = -.72), $t_{(45)} = -22.07$, $p < .0001$, to a strong RSF bias for mirror-reversed chimeras (M_{lat} ratio = +.59), $t_{(45)} = 15.22$, $p < .0001$ (see Figure 2). Perceptual asymmetries were less strongly influenced by orientation of the crying chimeras, with a moderate LSF bias for normal orientation chimeras ($M_{\text{lat ratio}} = -.42$), $t(45) = -7.84$, $p < .0001$, but nonsignificant asymmetry for mirror-reversed chimeras ($M_{\text{lat ratio}} = +.03$). Simple effects tests found that the orientation effect was nonetheless significant for both crying $F_{(1,45)} = 28.25$, $p < .0001$, and smiling chimeras, $F_{(1,45)} = 703.32$, $p < .0001$. Furthermore, the emotion effect was significant for both normal, $F_{(1,45)} = 24.14$, $p < .0001$, and mirror-reversed chimeras, $F_{(1,45)} = 63.29$, $p < .0001$. There were no significant sex effects or interactions.

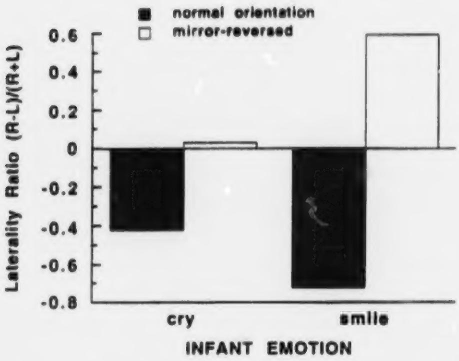


Figure 2. Interaction of infant emotion × chimera orientation in Experiment 2.

DISCUSSION

The main effect of emotion is consistent with our suggestion that valence effects may be optimized in perception of infant faces, perhaps due to increased emotional responsiveness to infants of the sort posited by ethological theory. Indeed, many subjects smiled or showed other positive emotional responses to the smiling infant faces, but none had done so while judging adult smiles in Experiment 1; conversely, crying infant faces often evoked sympathetic frowns or other emotional responses. Viewers showed a significant LSF bias in perception of negative infant expressions but no asymmetry for positive expressions. This pattern is most compatible with the negative-valence hypothesis of cerebral organization for emotional processes (e.g., Ehrlichman, 1988). The valence hypothesis (e.g., Silberman & Weingartner, 1986; Tucker, 1981) failed to find support for its prediction of a RSF-LH bias for positive expressions. Nor did the motivational hypothesis (e.g., Davidson, 1984) find support for the prediction that infant cries and smiles should yield a RSF-LH advantage because both infant expressions should elicit approach responses from adult viewers. The results also stand in contrast to the RH hypothesis' prediction that smiles should show the same overall LSF bias as cries.

In addition, the orientation effect shows a significant influence of infants' expressive asymmetries on adult perceptual field biases. Viewers showed a strong LSF bias for normal orientation chimeras, when infants' more expressive RF appeared on the left. But this shifted to a smaller yet significant RSF bias for mirror-reversed pairs, when infants' RF fell on the right.

Importantly, however, the significant interaction between emotion and orientation reveals that the relation between viewers' attentional biases and infant expressive asymmetries differed between judgments of negative and positive expressions. Although orientation (right/left position of infants' more expressive RF) influenced perception of both expressions, it did so differently for smiles and cries. The interaction pattern is reminiscent of the sex differences found for Test B in Experiment 1, and meets the negative-valence hypothesis' predictions of strong LSF bias for normal orientation cries, little or no bias for mirror-reversed cries, and an orientation-dependent shift in perceptual asymmetry concordant with the spatial hemifield containing the infants' more expressive RF. The obtained pattern was inconsistent with the predictions of each of the other three hypotheses.

Specifically, there was a trading relation between viewer attentional asymmetries and emitter expressive asymmetries in judgments of infant crying expressions, analogous to that for males' responses to *smiling* young men in Test B. Because the adult emitters' mean hemiface asymmetry was LF-biased while the infants' was RF-biased, however, viewer left hemifield attentional bias and emitter hemiface bias were concordant for *normal* orientation infant chimeras (as in face-to-face interactions) but discordant for *mirror-reversed* adult chimeras. Thus, the LSF bias was significant for normal-oriented infant cries, where attentional asymmetry and emitter asymmetry cooperate, but the two biases conflicted for mirror-reversed chimeras, resulting in a lack of perceptual asymmetry. In contrast, infant emitter asymmetries essentially overshadowed the impact of viewer attentional biases, analogous to the findings for females viewing smiling men in Test B. That is, judgments about intensity of infant *smiles* depended on which spatial hemifield contained the more expressive infant RH; they were influenced very little by viewers' attentional asymmetry. A strong LSF bias was found for normal-oriented infant smiles, but a strong RSF bias for mirror-reversed smiles. Recall that there were no sex effects in Experiment 2. Both male and female viewers showed this orientation by emotion interaction with infant faces, unlike the sex effect for adult faces in Test B.

Thus, Experiment 2 indicated a negative-valence effect on adult perceptual asymmetries for infant emotional expressions. However, it did not elucidate the perceptual processes underlying the phenomenon. One possibility is that negative expressions may be perceived as a configuration of the whole face (i.e., a gestalt of the combined features within the "frame" of face outline and hair), whereas perception of positive expressions may instead focus on the mouth as a singular distinguishing feature (Moscovitch, 1983). The holistic approach should call more heavily upon RH skills, while the feature-oriented approach should be better suited to LH analytic abilities (e.g., Levy, 1974; Bradshaw & Nettleton, 1981; Bryden, 1982; but see Trope, Rozin, Kemler Nelson & Gur, 1992). If the valence effect is attributable to such differences in perceptual approach to crying and smiling expressions, then the negative-valence effect—indeed the overall LSF bias—should become attenuated as the viewers' attention is progressively restricted to specific features of emotional information, such as the pattern of the central facial features taken out of their contextual "frame." This manipulation may lead subjects to use a more feature-oriented, analytic approach, and thus to rely more heavily on LH information processing strategies. Alternatively, the viewers' actual emotional responses to crying and smiling infants, rather than the information processing strategy, may be responsible for the valence effect. If so, the negative-valence effect should appear even when the viewer's attention is focused on facial-expression subcomponents or specific features. The next two experiments were designed to systematically examine these possibilities.

EXPERIMENT 3

If the holistic, gestalt-like perceptual specialization attributed to the RH accounts for the LSF for crying but not smiling infants, removal of the peripheral context such as a facial outline, cheeks and hair should attenuate or eliminate the negative-valence effect in perception of the remaining central facial features. To restrict the viewers' attention to the details of the central features of eyes/brows, mouth and nose, we deleted the unwanted peripheral "frame" information (i.e., face outline, ears, chin, cheeks, hair) by image-editing of optically-digitized versions of the original photographs, leaving only the facial features against a uniform white background. A new group of subjects made choices between pairs of the mixed-expression chimeras generated from these computer-edited expressions.

Method

Subjects. Ninety-six right-handed university and high school students (51 female, 45 male) participated.

Stimuli. High-quality photocopies of the original photographs from Experiment 2 were computer-digitized and edited, using an Apple Macintosh™ computer (see Best & Queen, 1989, for details). The cheeks, ears, chin, hair, and face outline were digitally erased from the digitized pictures, and the resulting images of the de-contextualized facial features were printed in normal and mirror-reversed orientation on white paper. Again obtaining judgments of mirror-image composites of each emitter's hemifaces, Best and Queen (1989) had found that these digitally-edited faces showed a strong right hemiface bias in expressiveness. These digitally-edited faces were used to generate mixed-expression chimeras (Figure 3),[2] which were assembled into a 40-page test booklet as before.

Procedure. Subjects completed the test booklet as in Experiment 2.

Results

Laterality ratios were analyzed as before. Significance for t-tests was again set at $p < .007$. There was a significant overall LSF bias ($M_{\text{lat ratio}} = -.11$), $t_{(95)} = -5.33$ $p < .0001$, the magnitude of which did not differ significantly from that in Experiment 2 according to t-test. The emotion effect was significant, $F_{(1,95)} = 7.76$, $p < .007$, again indicating a stronger LSF bias for crying ($M_{\text{lat ratio}} = -.15$), $t_{(95)} = -5.62$, $p < .0001$, than smiling expressions ($M_{\text{lat ratio}} = -.07$), $t_{(95)} = -3.10$, $p < .003$. The magnitude of the emotion difference in hemifield biases did not differ significantly from those in Experiment 2.

The orientation effect was significant, $F_{(1,95)} = 432.01$, $p < .0001$, indicating that infant expressive asymmetries affected viewers' judgments.

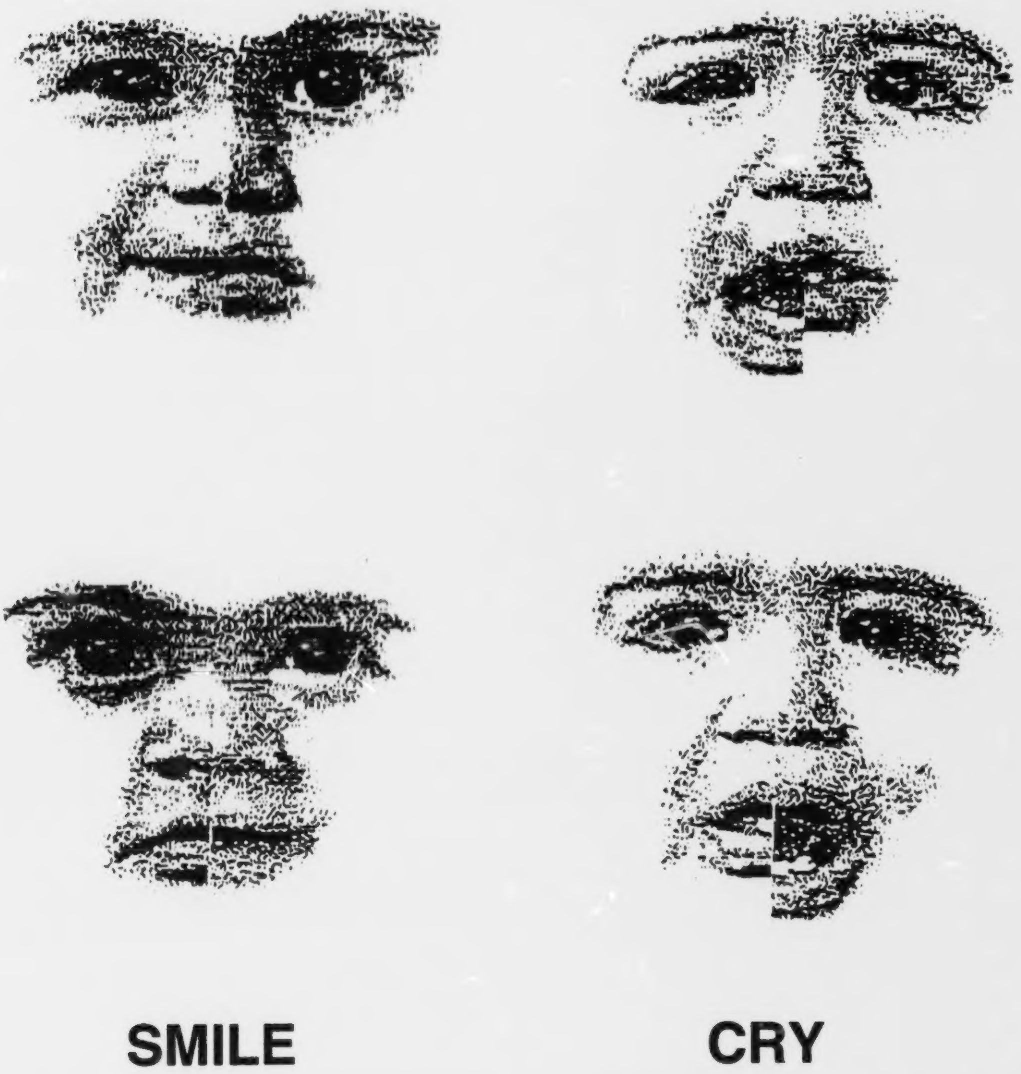


Figure 3. Examples of digitized mixed-expression chimeras of a smiling and a crying infant, with the emotional expressions on the left versus right side of the chimera, Experiment 3.

There was a significant LSF bias when the infants' more expressive RF was on the left in the normal orientation chimeras ($M_{\text{lat ratio}}$ = -.49), $t_{(95)}$ = -19.13, p < .0001, but a RSF bias when it was on the right in mirror-reversed chimeras ($M_{\text{lat ratio}}$ = +.26), $t_{(95)}$ = 8.83, p < .0001. The emotion × orientation interaction was also significant, $F_{(1,95)}$ < 63.64, p < .0001, repeating the pattern found in Experiment 2 (see Figure 4a). For smiling infants, normal orientation chimeras showed a strong LSF bias ($M_{\text{lat ratio}}$ = -.56), $t_{(95)}$ = -19.66, p < .0001, and mirror-reversed chimeras showed an equivalent, strong RSF bias ($M_{\text{lat ratio}}$ = +.42), $t_{(95)}$ = 12.34, p < .0001. The crying infants yielded a strong LSF bias for normal orientation chimeras ($M_{\text{lat ratio}}$ = -.41), $t_{(95)}$ = -11.08, p < .0001, but a smaller RSF bias for mirror-reversed chimeras ($M_{\text{lat ratio}}$ = +.11), $t_{(95)}$ = 2.94, p < .005.

Simple effect tests found the orientation effect to be significant for both crying, $F_{(1,95)} = 13.23$, $p < .0005$, and smiling, $F_{(1,95)} = 65.76$, $p < .0001$, and the emotion effect to be significant for both normal, $F_{(1,95)} = 559.26$, $p < .0001$, and mirror-reversed chimeras, $F_{(1,95)} = 104.62$, $p < .0001$.

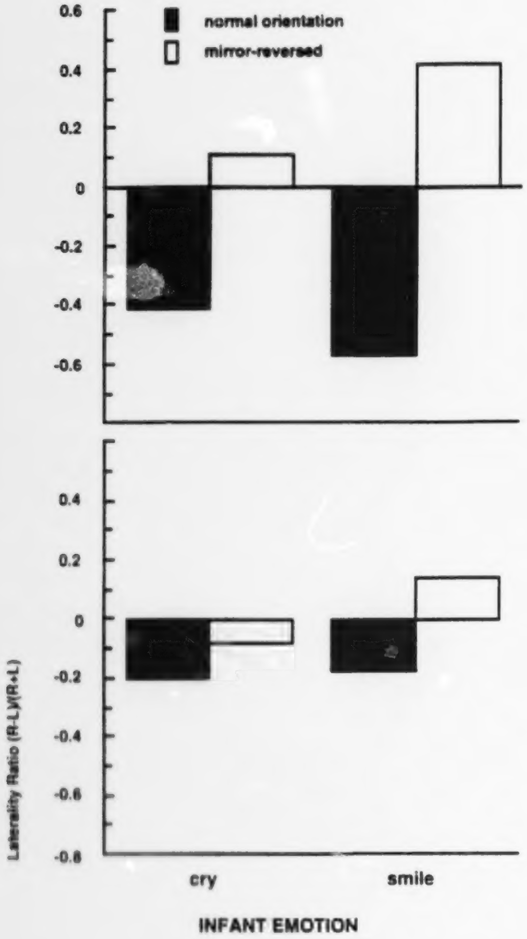


Figure 4. Interaction of infant emotion × chimera orientation in (a) Experiment 3 and (b) Experiment 4.

DISCUSSION

The results of this experiment replicated those of Experiment 2, even though the gestalt of the whole faces had been modified by removing the facial outline and other peripheral details, leaving only the central facial features. In fact, the magnitude of the effects failed to differ significantly from Experiment 2, suggesting that viewers' perception of the full-face chimeric photographs in the previous study had focused on the central facial features rather than their holistic relation to the contextual "frame" of facial outline, etc. It also suggests that the negative-valence effect may be due to some factor other than differential involvement of RH holistic and LH feature-analytic approaches to negative vs. positive expressions, respectively.

Perhaps, however, the stimulus manipulations of Experiment 3 failed to disrupt the facial gestalt sufficiently to interfere with a holistic RH response to crying expressions. The next experiment investigated this possibility by narrowing viewers' focus to specific facial features.

EXPERIMENT 4

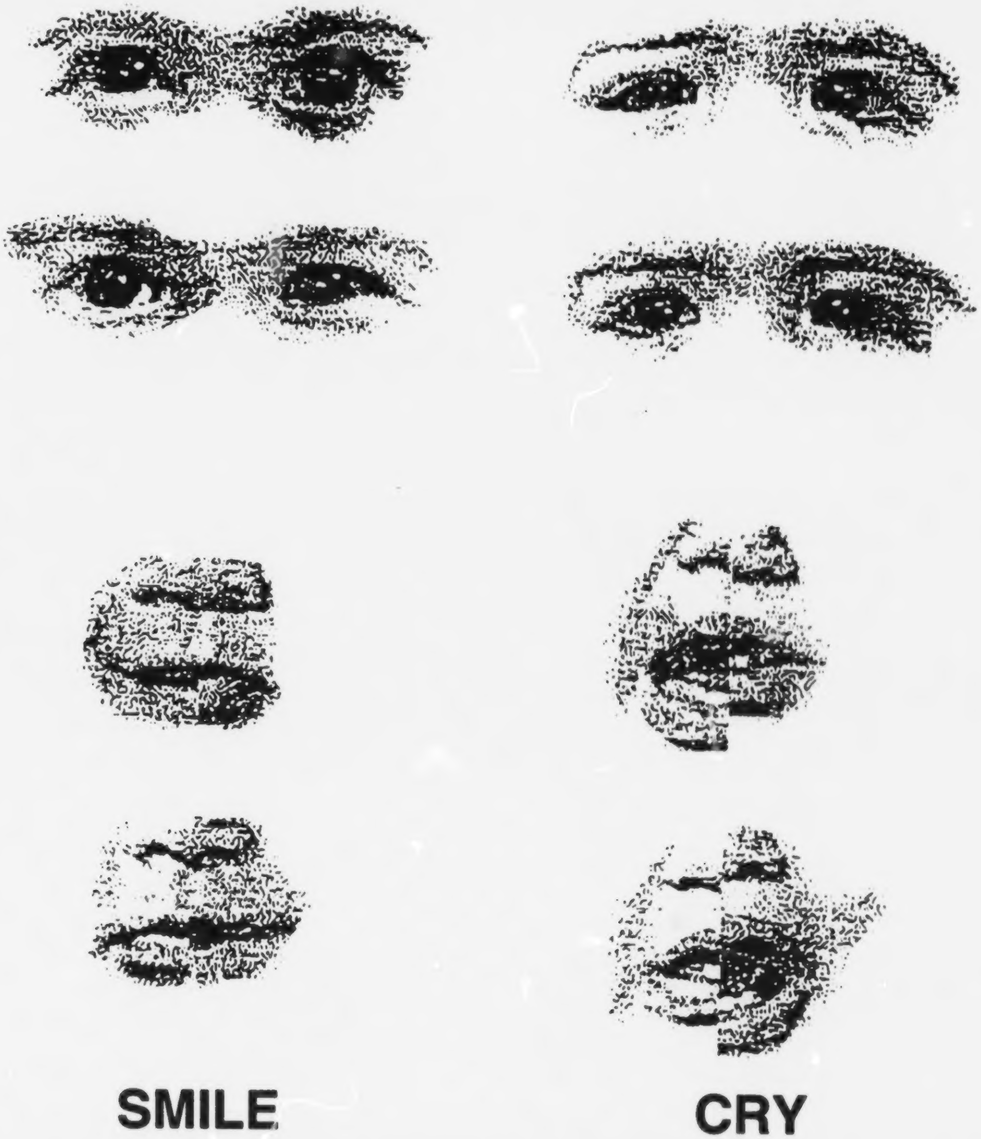
Restricting the view of infant faces to the mouth or eye/brow region alone should bias viewers' perceptual approach toward the analytic, feature-oriented abilities ascribed to the LH. If information processing differences between smiles and cries were responsible for the negative-valence effect as reasoned in Experiment 3, then this manipulation should either eliminate the valence effect or shift it to a strong RSF bias for smiles but a weak or nonexistent LSF bias for cries. However, if the negative-valence effect arises from emotional rather than cognitive factors, it should be impervious to this manipulation.

We focused on the expressive patterning of the mouth versus the eyes because our previous report (Best & Queen, 1989) had found that the infants' RF expressive bias was specific to the mouth, and was not present in the eye region; this eye/mouth asymmetry held for both smiles and cries. Viewers nonetheless were able to reliably judge relative happiness/sadness for either facial region. Each of these regions carries distinctive information in smiles and cries due to differential actions of the *zygomaticus, mentalis, levator palpebralis, orbicularis oculi*, and other facial muscles (Ekman, 1979; Oster & Ekman, 1978). Given that cortical input to the mouth region is contralateral, whereas input to the eye region is bilateral, our earlier results had suggested that lateralized cortical specializations rather than more peripheral factors are responsible for the RF bias in infant expressions. Thus, a second purpose of the present experiment was to test whether adults' perceptual asymmetries are influenced by the difference in asymmetrical patterning between the eye and the mouth regions of the infants' expressions. For Experiment 4, a new group of judges was presented with an "upper face" test and a "lower face" test employing further modifications of the digitized, edited infant expressions.

Method

Subjects. Participants were 54 right-handed university students (27 female, 27 male).

Stimuli. The digitized, edited faces from Experiment 3 were revised to produce an "upper face" test, for which all facial features other than the eyes, brows and bridge of the nose were removed , and a "lower face" test, for which all features other than the mouth and the tip of the nose were eliminated. Mixed-expression chimeras were generated separately for the eyes/brows and for the mouth (see examples in Figure 5, which uses the same infant emitters as in Figure 3). The eye and mouth regions were not separated from one another until after the midline had been traced as in Experiment 2. Two 40-page booklets were constructed as before, one for the "lower face" test and one the "upper face" test.



Figures 5. **Examples of digitized mixed-expression chimeras of the eye and mouth regions of a smiling and crying infant, with the emotional expressions on the left versus the right side of the chimera, Experiment 4.**

Procedure. The subjects were tested as before. Each completed the "lower face" test first and the "upper face" test second. Pilot testing had suggested that judgments about the eyes/brows might be more difficult than judgments about the mouth; this test order thus allowed more practice before the more difficult test.

Results

The data were handled as before, except that the ANOVA included a fourth factor: face part (mouth vs. eyes). Significance for multiple *t*-tests was set at $p < .002$.

Once again, there was an overall LSF bias (M_{lat} ratio = -.08), $t(53) = -4.44$, $p < .0001$, which did not differ significantly from the two preceding experiments. The emotion effect was again significant, $F_{(1,53)} = 13.96$, $p < .0005$. The facial regions from cries elicited a significant LSF bias (M_{lat} ratio = -.14, $t(53) = -5.14$, $p < .0001$, but those from smiles yielded no significant bias (M_{lat} ratio = -.02). The magnitude of this valence effect again failed to differ significantly from the earlier experiments. As before, there was also a significant orientation effect, $F(1,53) = 39.70$, $p < .0001$. The LSF bias held only for normal orientation, when the infants' RF was on the left side of the chimeras (M_{lat} ratio = -.19), $t(53) = -8.47$, $p < .0001$. The mirror-reversed chimeras elicited no significant bias (M_{lat} ratio = +.03). The emotion × orientation interaction was significant as well, $F(1,53) = 8.97$, $p < .004$ (see Figure 4b). As before, orientation had a smaller effect on perception of crying than smiling expressions. There was a LSF bias for normal orientation cries (M_{lat} ratio = -.20, $t(53) = -5.62$, $p < .0001$, but not for mirror-reversed cries (M_{lat} ratio = -.08). In contrast, normal orientation smiles evoked a LSF bias (M_{lat} ratio = -.17), $t(53) = -6.99$, $p < .0001$, but mirror-reversed smiles produced a significant RSF bias (M_{lat} ratio = +.14), $t(53) = 4.79$ $p < .0001$. Simple effects tests found the orientation effect to be significant for both crying, $F_{(1,53)} = 5.13$, $p = .02$, and smiling, $F_{(1,53)} = 76.97$, $p < .0001$. However, the emotion difference was significant only for mirror-reversed chimeras, $F(1,53) = 21.28$, $p < .0000$.

The face part factor also entered into two significant interactions. The face part × orientation interaction, $F(1,53) = 101.86$, $p < .0001$, showed that infant expressive asymmetries had a greater influence on perception of the mouth than the eyes. Normal orientation mouth chimeras yielded a LSF perceptual bias (M_{lat} ratio = -.35), $t(53) = -10.31$, $p < .0001$, while mirror-reversed mouths yielded a RSF bias (M_{lat} ratio = +.20), $t(53) = 4.93$, $p < .0001$. In contrast, the eyes produced a smaller but significant LSF bias in *mirror-reversed* orientation (M_{lat} ratio = -.14), $t(53) = -3.98$, $p < .0002$, which became nonsignificant for normal orientation (M_{lat} ratio = -.03). Simple effects tests found that the face part difference was significant for both normal orientation, $F_{(1,53)} = 62.39$, $p < .0001$, and mirror-reversed items, $F_{(1,53)} = 44.88$, $p < .0001$. Moreover, the orientation effect was significant both for the mouth, $F_{(1,35)} = 119.58$, $p < .0001$, and for the eyes, $F_{(1,53)} = 6.68$, $p < .01$.

The face part × orientation × emotion interaction was also significant, $F_{(1,53)} = 165.97$, $p < .0001$. As with the full-face studies, smiling mouths produced a large LSF bias for normal orientation (M_{lat} ratio = -.57), $t(53) = -17.31$, $p < .0001$, and a large RSF bias for mirror-reversed items (M_{lat} ratio = +.52), $t(53) = 11.92$, $p < .0001$. Crying mouths showed nonsignificant LSF biases for normal orientation (M_{lat} ratio = -.13), and mirror-reversed chimeras (M_{lat} ratio = -.11). Crying eyes yielded a significant LSF bias for normal orientation chimeras (M_{lat} ratio = -.28, $t(53) = -6.23$, $p < .0001$, but no bias for mirror-reversed ones (M_{lat} ratio = -.05), consistent with previous full-face results. Smiling eyes, however, elicited a modest RSF bias for normal orientation (M_{lat} ratio = +.23), $t(53) = 5.67$ $p < .0001$, and an equal LSF bias for mirror-reversed chimeras (M_{lat} ratio = -.24), $t(53) = -4.95$, $p < .0001$. The direction of this orientation effect for smiling eyes was opposite that of the emotion × orientation interactions found in Experiments 2 and 3, where the normal orientation was associated with LSF bias and the mirror-reversed with RSF bias.

Overall, then, orientation again had a greater effect on perceptual responses toward the smiling expressions than toward the crying expressions. According to simple effects tests, the orientation effect was significant for crying eyes, $F_{(1,53)} = 53.48$, $p < .0006$, for smiling mouths, $F_{(1,53)} = 457.26$, $p <.0001$, and for smiling eyes, $F_{(1,53)} = 11.06$, $p < .002$, but not for crying mouths. The emotion effect was significant for eyes in both normal, $F_{(1,53)} = 56.96$, $p < .0001$, and mirror-reversed orientation, $F_{(1,53)} = 9.72$, $p < .003$, as well as for mouths in both normal, $F_{(1,53)} = 58.74$, $p < .0001$, and mirror-reversed orientation, $F_{(1,53)} = 90.83$, $p < .0001$.

DISCUSSION

The emotion main effect was not diminished relative to the two other infant face tests, in spite of restricting the viewers' attention to isolated facial features. This finding suggests that the negative-valence effect on perceptual asymmetries for infant emotional expressions derives from emotional processes rather than information processing factors. The emotion × orientation interaction again indicated that, overall, there was a weaker hemiface effect, or greater effect of attentional asymmetry, on perception of crying than smiling expressions. Moreover, differences in perception of the eye and mouth regions suggest that the viewers were sensitive to differences in the expressive asymmetries displayed by those facial regions. Consistent with the Best & Queen (1989) finding of a significant overall RF bias only for the mouth region, the viewers in the present study were more affected overall by orientation of the mouth than of the eyes. We should note, however, that this face part interaction differed for smiles versus cries. For smiles, both face regions showed dominance of the orientation factor, as before, but the direction of this influence was reversed for the eyes relative to the mouth and to the previous studies. That is, viewers apparently detected greater intensity of expression on the *left* hemiface for smiling eyes, but on the *right* for the mouth. For crying expressions, there was a greater perceptual effect of orientation on eyes than mouth. The pattern of higher-order face part effects are curious, given the Best and Queen finding (1989) that only mouths showed significant RF expressive asymmetry. Although a complete explanation cannot be offered at this time, this interaction nonetheless indicates that adults are quite sensitive to emotional information in the eye region of infant expressions.

To summarize, the perceptual findings with infant smiles and cries in Experiments 2-4 provided fairly strong support for the negative-valence hypothesis over the other three hypotheses. However, those results were based on the same set of six smiling and four crying infants. Therefore, it was important to extend our investigation to a new set of infant photographs.

EXPERIMENT 5

In the three preceding studies the smiling and crying expressions had come from different infants. Although the mean rated intensities (Hildebrandt, 1983) of the two types of expression were roughly equivalent, they were not absolutely matched. These factors left open the (unlikely, we thought) possibility that individual differences in infant expressiveness and/or differences in the mean intensity of the two expression types might account for the main effects of emotion found in Experiments 2 and 3, or for the pattern of the emotion by orientation interactions.

We wished to insure that the infants' expressions of happiness and distress were spontaneous and genuine. Although the laboratory photographs of infants from Hildebrandt and Fitzgerald (1978, 1979, 1981) seemed appropriate for our purposes, based on reports that infants do not mask or simulate emotions until their second year (Campos et al., 1983; Oster & Ekman, 1978; Rothbart & Posner, 1985; Sroufe, 1979), a recent study suggests that infants *do* produce smiles like those of adults simulating happiness they don't feel or covering up negative emotions. Ekman & Friesen (1982) termed such adult expressions "unfelt smiles," as described earlier. Those authors claim that whereas felt smiles are virtually symmetrical, unfelt smiles tend to show asymmetries favoring the LF (recall, however, the difficulties presented to this position by the asymmetrical adult smiles in the Levy et al. stimuli used in Experiment 1). Fox and Davidson (1988) videotaped infants responding to mother versus a stranger in an unfamiliar laboratory setting, and found evidence of unfelt smiles, i.e., lacking *orbicularis oculi* activity, toward strangers but not toward mother. Moreover, EEG asymmetry patterns over the infants' frontal lobes differed between felt (LH activation) and unfelt smiles (RH activation). The facial expressions we used in Experiments 2-4 had been obtained by a portrait photographer (i.e., a stranger) in a university laboratory (i.e., unfamiliar setting). Both factors raise the likelihood of unfelt smiles, and the possibility that the RF bias we had found in those smiles might not occur in genuine, felt smiles. Indeed, as mentioned earlier, only two of those six smiling expressions showed evidence of *orbicularis oculi* activity.

For these reasons, one of us (JSW) photographed infants enrolled in high-quality daycare, a familiar and comfortable setting to the infants. During a two-month period JSW visited the daycare centers 2-3 days per week to interact with the infants. Before she began to photograph the infants at a given center, she spent at least three weeks there, playing with the infants, interacting with their caregivers, and participating in daily caregiving (e.g., feeding, diaper-changing). Thus, she was not a stranger but had become a familiar caregiver. After she had become familiar to the infants, she took multiple photographs of each infants' expressions, taking care to "catch them in the act" of spontaneous social smiles and distress cries, as well as of neutral expressions. We selected for this study only those emitters for whom a smile and a cry photo were matched in emotional intensity, according to a preliminary rating study. These photographs were then used to assess infants' expressive asymmetries for spontaneous smiles and cries, as well as to extend the investigation of perceptual asymmetry. This study was modeled after Experiment 2, using photographs rather than digitally-edited images. There had been remarkable consistency in the major findings of the preceding studies, indicating that the primary effects had not been influenced substantially by the progressive restriction of facial features available for judgments. Therefore, we used the full facial configurations of the actual photographs in the present study.

Method

Subjects. Forty-four right-handed university students (22 male, 22 female) participated. Five more failed the handedness criteria (n=3) or filled out their answer sheets incorrectly (n=2).

Stimuli. The spontaneous neutral, crying, and smiling expressions of 17 infants (range = 5-14 mo.) were photographed at their daycare, using black and white print film in a Minolta XG-1 camera fitted with a zoom lens. All were printed, placed behind an oval template as in Experiment 2 to screen out peripheral and background details, and photocopied via a Xerox 1012 machine with a grayscale setting. The first 12 infants provided 68 photos, which were compiled randomly into a pre-test intensity rating booklet. Twelve university students rated each expression between -3 (very sad) to +3 (very happy), with 0 as neutral. Nine infants had at least one smile and one cry that were rated equally intense (e.g., +2 and -2, respectively), along with one clearly neutral expression (rated 0). Therefore, five additional infants were photographed and their expressions submitted to 12 new raters; the latter infants all met the equal-intensity criterion. Of the final 14 infants, 12 showed clear *orbicularis occuli* activity (AU6-7), suggesting "felt" smiles.

Mixed-expression chimeras were constructed for the 14 infants with matched intensity and paired as before, for the first 56 pages of a new test booklet. Mirror-image composites of each hemiface for each infant's smile and cry were also constructed, as in Best and Queen (1989), to test for expressive asymmetries in the booklet's last 28 pages. Top-bottom position of right vs. left composites was counterbalanced over infants and expressions.

Procedure. For the first part of the test, subjects judged mixed-expression chimeras. For the second part, they judged left vs. right mirror-composites of each expression for each infant.

Results

Mirror-image composites. Because the interpretation of orientation effects on judgments of mixed-expression chimeras depends on the expressive asymmetries observed in the emitters, we begin by reporting on the test for hemiface biases in infant smiles and cries. Laterality ratios were computed on choices of the left vs. right hemiface mirror-composites and analyzed in a 2 × 2 ANOVA (sex of viewers × infant emotion). Alpha level for t-tests was set at $p < .025$.

Only the main effect of emotion was significant, $F_{(1,42)} = 34.931$, $p < .0001$, reflecting a right hemiface bias in intensity of crying expressions ($M_{\text{lat rat}} = +.263$), $t_{(43)} = 8.736$, $p < .0001$, but a nonsignificant *left*-side bias in smiles ($M_{\text{lat rat}} =$ -.013). That is, these spontaneous smiles failed to show the rightward bias found in previous reports (Best & Queen, 1989; Rothbart et al., 1989) and in the smiling expressions used in Experiments 2-4, although crying expressions replicated the earlier-found right hemiface bias. Thus, the new set of mixed-expression chimeras were expected to yield the same orientation effect for crying chimeras as found before, but there should be *no* orientation difference for smiling chimeras, unlike Experiments 2-4.

Mixed-expression chimeras. Laterality ratios were entered into a 2 × 2 × 2 ANOVA (sex × emotion × orientation). The alpha criterion for multiple t-tests was set at $p < .00625$.

There was a significant LSF bias overall ($M_{\text{lat rat}} = -.312$), $t_{(43)} = 6.92$, $p < .0001$. However, the significant effect of emotion, $F_{(1,42)} = 32.16$, $p <$ .0009, indicates that the LSF bias for smiles ($M_{\text{lat rat}} = -.44$), $t_{(43)} = -7.78$, $p < .0001$, was larger than that for cries ($M_{\text{lat rat}} = -.18$), although cries were significantly LSF-biased, $t_{(43)} = -3.84$, $p < .0004$. The emotion × orientation interaction was also significant, $F_{(1,42)} = 6.754$, $p < .0129$ (see Figure 6). Simple effect tests found significant orientation effects for both cries, $F_{(1,42)} = 17.188$, $p < .001$, and smiles, $F_{(1,42)} = 23.347$, $p < .001$. The difference between expressions was significant for normal, $F_{(1,42)} = 80.152$, $p < .0001$, but not mirror-reversed orientation. An LSF bias appeared for smiles in normal ($M_{\text{lat rat}} = -.31$), $t_{(43)} = -4.69$, $p < .0001$, and mirror-reversed orientation ($M_{\text{lat rat}} = -.57$), $t_{(43)} = -9.61$, $p <$

.0001, and for cries in normal orientation ($M_{\text{lat rat}}$ = -.28), $t_{(43)} = -7.78$, $p < .0001$. As in Experiments 2-4, perceptual asymmetry was lacking for mirror-reversed cries ($M_{\text{lat rat}}$ = -.08).

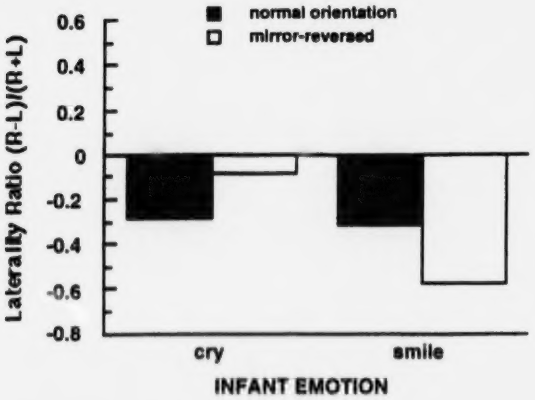


Figure 6. Interaction of infant emotion × chimera orientation on the mixed-expression chimera trials in Experiment 5.

DISCUSSION

The mirror-image composites revealed a significant RF expressive bias for infant crying, consistent with previous reports (Best & Queen, 1989; Rothbart et al., 1989). However, these spontaneous smiles showed no asymmetry. Thus, spontaneous infant smiles and cries show expressive asymmetries that, like the perceptual results of Experiments 2-4, support the negative-valence hypothesis. That these smiles were symmetrical, while RF bias is reported for smiles obtained under conditions that may foster "unfelt" or socially conditioned expressions (see Fox & Davidson, 1988), is also compatible with claims (Ekman & Friesen, 1979) that truly spontaneous, genuine smiles fail to show significant asymmetry.

Given the replicated RF bias for cries, the same interaction of emitter hemiface bias and viewer attentional bias on perception of mixed-expression crying chimeras should occur as in Experiments 2-4. This was exactly the result obtained. As before, an LSF bias in perception occurred for normal-orientation cries, where the more expressive infant RF fell in the viewer's left spatial hemifield, but disappeared for mirror-reversed cries, where the RF fell in the *less* sensitive hemifield. However, the near-symmetry of spontaneous smiles in this experiment substantially changed the perceptual pattern found for smiles in Experiments 2-4, which was essentially determined by which hemifield contained the infants' more expressive RF. Specifically, this time *both* chimera orientations yielded an LSF bias for infant smiles. That is, when the hemiface bias of the smiles is extremely weak, it no longer dominates the viewer's perceptual asymmetry. Instead, an underlying leftward attentional bias appears, as was found for adult smiles in Test A of Experiment 1, where hemiface biases were eliminated by the pairing of chimeras. Nonetheless, the present finding for spontaneous infant smiles still differed in an important way from the Test A pattern. Remarkably, these very weakly asymmetrical infant smiles *still* produced a significant orientation effect on *degree* of LSF bias. The tiny, nonsignificant LF bias in spontaneous infant smiles produced a significantly larger viewer LSF bias when the infant LF appeared in the more sensitive left hemifield for mirror-reversed pairs than when it appeared in the *less* sensitive right hemifield for normal orientation pairs.

GENERAL DISCUSSION

Taken together, the results indicate that the relative contributions of viewer attentional bias and emitter expressive bias on adult judgments of infant emotional expressions differ for crying and smiling. The perceptual asymmetries as well as hemiface biases in infants' spontaneous expressions (Experiment 5) both support the negative-valence model of emotional asymmetries (e.g., Dimond, et al., 1976; Ehrlichman, 1988; Sackeim & Gur, 1978, 1980). At least when viewing static infant faces, adults show a RH bias for negative emotion, which interacts with asymmetries in the infants' faces. But adults' perception of positive emotion in infants is dominated by asymmetry in the expressions, which overpowers adults' attentional bias toward the LSF *unless* the expressive asymmetry is very weak, as in spontaneous smiles.

The other three models of emotional asymmetry did not fare as well. The RH model predicts the same perceptual pattern for negative and positive emotions, yet there were significant differences. The valence hypothesis posits RH specialization for negative emotion and LH specialization for positive emotion; however, perception of infant smiles failed to show an overall RSF-LH bias, and their spontaneous smiles showed no expressive asymmetry. As for the motivational hypothesis, an RSF/LH bias should result from approach responses to infant smiles, as we argued for cries also, yet neither showed that perceptual bias. It should be noted, however, that a more stringent test of the motivational hypothesis would require direct assessment of viewers' motivational tendencies toward the infant emitters.

Given that the majority of findings on perceptual asymmetries for *adult* facial expressions have supported the RH hypothesis, the present infant face findings suggest that valence effects on perceptual asymmetries may depend on viewers' emotional responses. Although this and other tasks supporting valence effects have called for emotionality judgments, that alone may not suffice to produce a negative-valence effect on perception. Levy et al. (1983) and Experiment 1 required emotionality judgments about adult chimeras, yet those studies found a significant LSF-RH bias for smiles. Infant smiles, which should increase viewers' emotional responses, instead yielded no overall perceptual asymmetry (Experiments 2, 4, 5) or at best a small LSF bias (Experiment 3). A separate study from our lab provided additional corroboration of perceptual differences for infant versus adult expressions. Chaiken (1988) employed two chimeric choice tasks with 7-15 year old children and adults, using both adult and infant expressions, and found a valence effect only for infant expressions. However, viewer emotional responses will need to be assessed directly in future studies to test whether this factor is indeed crucial to a valence effect on perception. Such information may be especially critical for more comprehensive test of the motivational model than provided in the studies reported here.

Experiments 3 and 4 suggest that the negative-valence effect on perception was not due to basic information processing differences between the hemispheres for negative versus positive expressions. Manipulations designed to restrict viewers' attention to progressively narrower features of the infant expressions should have shifted perception toward the analytical, feature-oriented approach of the LH, yet did not influence overall perceptual asymmetry. Nor, more importantly, did they change the valence effect. Thus, the negative-valence effect for infant expressions seems to reflect an aspect of hemispheric specialization that is largely independent of information processing asymmetries.

As noted earlier, adult's LSF-RH bias in perception of infants' crying expressions is compatible with the greater intensity of expressions on the infant's RF (Best & Queen, 1989; Rothbart, Taylor & Tucker, 1989). In face-to-face interactions, the infant's more expressive hemiface appears in the adult's more sensitive LSF, presumably enhancing the adult's emotional response. This compatibility does not hold in the case of adult face-to-face interactions, given that adults show a *LF* expressive bias, which falls in the viewer's *less* sensitive RSF. Generally enhanced sensitivity and responsiveness to infant expressions is consistent with ethological theory. But why should the interaction between infant expressive asymmetry and adult attentional bias differ between crying and smiling expressions? Perhaps it can be related to differences in the imperativeness of adult responses to infant distress and pleasure states. Presumably, infant distress indicates a possible danger to the infant, or some health or survival need, which would impel caregivers or other adults to take action on the infant's behalf. In contrast, an infant's smile does not signal this sort of urgency. Therefore, the evolutionary pressure for specialized responsiveness toward infant crying expressions may have been greater than, or at least qualitatively different from, that toward infant smiles. Specialized responsiveness to infant cries may be optimized by the interaction between infant expressive asymmetries and the viewer's LSF attentional bias, which may provide for the most direct, immediate activation of the RH motivational and/or action systems that are specialized for rapid responses to potentially threatening situations. The notion that the right hemisphere is specialized for response to affectively negative situations that mobilize fleeing behavior (rapid withdrawal) was proposed by Kinsbourne (1978) and further developed by Davidson (1984). Supporting evidence has been found in infants' EEG asymmetries during facial expressions of distress in response to stranger approach and maternal separation, as well as during newborns' facial disgust responses to noxious gustatory stimuli (Fox & Davidson, 1986, 1987, 1989). Moreover, an evolutionary foundation for this bias is suggested by two recent studies of monkeys. In one, rhesus monkeys displayed earlier-appearing and more intense negative emotional expressions on the left (RH) than the right hemiface (Hauser, 1993). In the other report, which is particularly germane to the present argument, rhesus mothers consistently picked up their infants with the left hand when frightened by the approach of a human (Haida & Koichi, 1991), but used either hand in neutral situations.

REFERENCES

Ahern, G. L., & Schwartz, G. E. (1979). Differential lateralization for positive versus negative emotion. *Neuropsychologia, 17,* 693-698.

Best, C. T., & Queen, H. F. (1989). Baby, it's in your smile: Right hemiface bias in infant emotional expressions. *Developmental Psychology, 25,* 264-276.

Bowlby, J. (1969). *Attachment and loss (Vol. 1).* New York, NY: Basic Books.

Bradshaw, J. L., & Nettleton, N. C. (1981). The nature of hemispheric specialization in man. *Behavioral and Brain Sciences, 4,* 51-91.

Brody, N., Goodman, S. E., Halm, E., Krinzman, S., & Sebrechts, M. M. (1987). Lateralized affective priming of lateralized affectively valued target words. *Neuropsychologia, 25,* 935-946.

Bryden, M. P. (1982). *Laterality: Functional asymmetry in the intact brain.* New York: Academic Press.

Bryden, M. P., & Ley, R. G. (1983). Right hemispheric involvement in imagery and affect. in E. Perecman (ed.) *Cognitive processing in the right hemisphere.* New York: Academic Press.

Bryden, M. P., & Sprott, D. A. (1981). Statistical determination of degree of laterality. *Neuropsychologia, 19,* 571-581.

Buck, R. (1986). The psychology of emotion. In J. E. LeDoux & W. Hirst (Eds.), *Mind and brain: Dialogues in cognitive neuroscience* (pp. 275-300). London, England: Cambridge University Press.

Burton, L. A., & Levy, J. (1991). Effects of processing speed on cerebral asymmetry for left- and right-oriented faces. *Brain and Cognition, 15,* 95-105.

Campbell, R. (1978). Asymmetries in interpreting and expressing a posed facial expression. *Cortex, 19,* 327-342.

Campos, J. J., Barrett, K. C., Lamb, M. E., Goldsmith, H. H., & Stenberg, C. (1983). Socioemotional development. In P. H. Mussen (ed.), M. M. Haith & J. J. Campos (Vol. eds.), *Handbook of child development: Vol. II. Infancy and developmental psychobiology* (pp. 784-915). New York: John Wiley & Sons.

Carlson, D. F., & Harris, L. J. (1985). Perception of positive and negative emotion in free viewing of asymmetrical faces: Sex and handedness effects. Paper presented at the *International Neuropsychology Society* meeting, February 6-8, San Diego.

Chaiken, S. (1988). *A free vision test of developmental changes in hemispheric asymmetries for perception of infant and adult emotional expressions.* Unpublished M. A. thesis, Wesleyan University, Middletown, CT.

Chaurasia, B. D., & Goswami, H. K. (1975). Functional asymmetry in the face. *Acta Anatomica, 91,* 154-160.

Davidson, R. J. (1984). Affect, cognition, and hemispheric specialization. In C. E. Izard, J. Kagan, & R. Zajonc (Eds.), *Emotion, cognition and behavior* (pp. 320-365). New York: Cambridge University Press.

Davidson, R. J. (1992). Anterior cerebral asymmetry and the nature of emotion. *Brain and Cognition, 20,* 125-151.

Davidson, R. J., & Fox, N. A. (1982). Asymmetrical brain activity discriminates between positive and negative affective stimuli in human infants. *Science, 218,* 1235-1237.

Davidson, R. J., Schwartz, G. E., Saron, C., Bennett, J., & Goleman, D. J. (1979). Frontal versus parietal EEG asymmetry during positive and negative affect. *Psychophysiology, 16,* 202-203.

DeRenzi, E., Gentillini, M., Faglioni, P. Barbieri, C. (1989). Attentional shift towards the rightmost stimuli in patients with left visual neglect. *Cortex, 25,* 231-237.

Dimond, S. J., & Farrington, L. (1977). Emotional response to films shown to the right and left hemisphere of the brain measured by heart rate. *Acta Psychologica, 41,* 255-260.

Dimond, S. J., Farrington, L., & Johnson, P. (1976). Differing emotional response from right and left hemispheres. *Nature, 261,* 690-692.

Ehrlichman, H. (1988). Hemispheric asymmetry and positive-negative affect. In D. Ottoson (ed.), *Duality and the unity of the brain.* London: MacMillan Press.

Eibl-Eibesfeldt, I. (1975). *Ethology: The biology of behavior.* New York: Holt, Rhinehart and Winston.

Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. *Nebraska symposium on motivation* (pp. 207-283). Lincoln: University of Nebraska Press.

Ekman, P. (1979). About brows: Emotional and conversational signals. In M. von Cranach, K. Foppa, W. Lepenies, and D. Ploog (Eds.) *Human ethology.* pp. 169-249. London: Cambridge University Press.

Ekman, P., & Freisen, W. V. (1978). *Facial action coding system.* Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., & Friesen, W. V. (1982). Felt, false, and miserable smiles. *Journal of Nonverbal Behavior, 6,* 238-252.

Emde, R. N., Gaensbauer, T. J., & Harmon, R. J. (1976). *Emotional expression in infancy: A biobehavioral study.* New York: International Universities Press.

Fox, N. A., & Davidson, R. J. (1986). Taste-elicited changes in facial signs of emotion and the asymmetry of brain electrical activity in newborn infants. *Neuropsychologia, 24,* 417-422.

Fox, N. A., & Davidson, R. J. (1987). Electroencephalogram asymmetry in response to the approach of a stranger and maternal separation in 10-month-old infants. *Developmental Psychology, 23,* 233-240.

Fox, N. A., & Davidson, R. J. (1988). Patterns of brain electrical activity during facial signs of emotion in 10-month-olds. *Developmental Psychology, 24,* 230-236.

Gage, D. F., & Safer, M. A. (1985). Hemisphere differences in the mood state-dependent effect for recognition of emotional faces. *Journal of Experimental Psychology: Learning, Memory and Cognition, 11,* 752-763.

Gainotti, G. (1972). Emotional behavior and hemispheric side of lesion. Cortex, 8, 41-55.

Gainotti, G. (1988). Disorders of emotional behavior and of autonomic arousal resulting from unilateral brain damage. In D. Ottoson (Ed.), *Duality and the unity of the brain.* London: MacMillan Press.

Green, J., Morris, R. D., Epstein, C. M., West, P. D., & Engler, H. F. (1992). Assessment of the relationship of cerebral arousal asymmetry to perceptual asymmetry. *Brain and Cognition, 20,* 264-279.

Haida, T., & Koichi, M. (1991). Left-hand preference in frightened mother monkeys in taking up their babies. *Neuropsychologia, 29,* 207-209.

Hauser, M. (1993). Right hemisphere dominance for the production of facial expression in monkeys. *Science, 261,* 475-477.

Heller, W., & Levy, J. (1981). Perception and expression of emotion in right-handers and left-handers. *Neuropsychologia, 19,* 263-272.

Hellige, J. B., Bloch, M.. I., Taylor, A. K. (1988). Multitask investigation of individual differences in hemispheric asymmetry. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 178-187.

Hildebrandt, K. A. (1983). Effect of facial expression variations on ratings of infants' physical attractiveness. *Developmental Psychology, 19,* 414-417.

Hildebrandt, K. A., & Fitzgerald, H. E. (1978). Adults' responses to infants varying in perceived cuteness. *Behavioural Processes, 3,* 159-172.

Hildebrandt, K. A., & Fitzgerald, H. E. (1979). Adults' perceptions of infant sex and cuteness. *Sex Roles, 5,* 471-481.

Hildebrandt, K. A., & Fitzgerald, H. E. (1981). Mothers' reponses to infant physical appearance. *Infant Mental Health Journal, 2,* 56-61.

Hirschman, R. S., & Safer, M. A. (1982). Hemisphere differences in perceiving positive and negative emotions. *Cortex, 18*, 569-580.

Izard, C. (1979). *The maximally discriminative facial movement coding system.* Newark: University of Delaware Press.

Izard, C., Huebner, R. R., Risser, D., McGinnes, G. C., & Dougherty, L. M. (1980). The young infant's ability to produce discrete emotional expressions. *Developmental Psychology, 16*, 132-140.

Kim, H., Levine, S. C., & Kertesz, S. (1990). Are variations among subjects in lateral asymmetry real individual differences or random error in measurement? Putting variability in its place. *Brain and Cognition, 14*, 220-242.

Kinsbourne, M. (1970). The cerebral basis of lateral asymmetries in attention. *Acta Psychologica, 33*, 193-201.

Kinsbourne, M. (1978). The biological determinants of functional bisymmetry and asymmetry. In M. Kinsbourne (Ed.) *Asymmetrical functions of the brain.* New York: Cambridge University Press.

Kuhn, G. M. (1973). The Phi coefficient as an index of ear differences in dichotic listening. *Cortex, 9*, 447-457.

Ladavas, E., Umilta, C., & Ricci-Bitti, P. E. (1980). Evidence for sex differences in right hemisphere dominance for emotions. *Neuropsychologia, 18*, 361-366.

Levine, S. C., & Levy, J. (1986). Perceptual asymmetry for chimeric faces across the life span. *Brain and Cognition, 5*, 291-306.

Levy, J. (1974). Psychobiological implications of bilateral asymmetry. In S. J. Dimond & G. Beaumont (Eds.) *Hemisphere function in the human brain.* London: Elek Books.

Levy, J., Heller, W., Banich, M. T., & Burton, L. A. (1983a). Asymmetry of perception in free viewing of chimeric faces. *Brain and Cognition, 2*, 404-419.

Levy, J., Heller, W., Banich, M. T., & Burton, L. A. (1983b). Are variations among right-handed individuals in perceptual asymmetries caused by characteristic arousal differences between the hemispheres? *Journal of Experimental Psychology: Human Perception and Performance, 9*, 329-359.

Levy, J., & Kueck, L. (1986). A right hemispatial field advantage on a verbal free-vision task. *Brain and Language, 27*, 24-37.

Levy, J., Wagner, N., & Luh, K. (1990). The previous visual field: Effects of lateralization and response accuracy on current performance. *Neuropsychologia, 28*, 1239-1249.

Ley, R. G, & Bryden, M. P. (1979). Hemispheric differences in recognizing faces and emotions. *Brain and Language, 7*, 127-138.

Ley, R. G., & Bryden, M. P. (1981). Consciousness, emotion, and the right hemisphere. In R. Stevens & G. Underwood (Eds.), *Aspects of consciousness: Structural issues.* New York: Academic Press.

Lorenz, K. Z. (1935). *Studies in animal and human behavior. Vol. 1.* Cambridge: Harvard University Press.

Lorenz, K. Z. (1981). *The foundations of ethology.* New York: Springer-Verlag.

Lorenz, K. Z., & Leyhausen, P. (1973). *Motivation of human and animal behavior.* New York: Van Nostrand.

Luh, K., Rueckert, L. M., & Levy, J. (1991). Perceptual asymmetries for free viewing of several types of chimeric stimuli. *Brain and Cognition, 16*, 83-103.

Mondor, T. A., & Bryden, M. P. (1992). On the relation between auditory spatial attention and auditory perceptual asymmetries. *Perception & Psychophysics, 52*, 393-402.

Monrad-Krohn, G. H. (1924). On the dissociation of voluntary and emotional innervation in facial paralysis of central origin. *Brain, 47*, 22-35.

Moscovitch, M. (1983). The linguistic and emotional functions of the normal right hemisphere. In E. Perecman (Ed.), *Cognitive processing in the right hemisphere.* New York: Academic Press.

Natale, M., Gur, R. E., & Gur, R. C. (1983). Hemispheric asymmetries in processing emotional expressions. *Neuropsychologia, 21*, 555-565.

Oster, H., & Ekman, P. (1978). Facial behavior in child development. In W. A. Collins (Ed.), *Minnesota Symposium on child psychology* (Vol. 2, pp. 231-276). Hillsdale, NJ: Lawrence Erlbaum Associates.

Oster, H., Hegley, D., & Nagel, L. (1992). Adult judgments and fine-grained analysis of infant facial expressions: Testing the validity of a priori coding formulas. *Developmental Psychology, 28*, 1115-1131.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13*, 25-42.

Remillard, G. M., Anderman, F., Rhi-Sausi, A., & Robbins, N. M. (1977). Facial asymmetry in patients with temporal lobe epilepsy: A clinical sign useful in the lateralization of temporal epileptic foci. *Neurology, 27*, 109-114.

Reuter-Lorenz, P., & Davidson, R. J. (1981). Differential contributions of the two cerebral hemispheres to the perception of happy and sad faces. *Neuropsychologia, 19*, 609-613.

Reuter-Lorenz, P. A., Givis, R. P., & Moscovitch, M. (1983). Hemispheric specialization and the perception of emotion: Evidence from right-handers and from inverted and non-inverted left-handers. *Neuropsychologia, 21*, 687-692.

Reuter-Lorenz, P. A., Kinsbourne, M., & Moscovitch, M. (1990). Hemispheric control of spatial attention. *Brain and Cognition, 12*, 240-266.

Rinn, W. E. (1984). The neuropsychology of facial expression: A review of the neurological and psychological mechanisms for producing facial expressions. *Psychological Bulletin, 95*, 52-77.

Rossi, G. F., & Rosadini, G. (1967). Experimental analysis of cerebral dominance in man. In C. H. Millikan & F. L. Darley (Eds.), *Brain mechanisms underlying speech and language.* New York: Grune & Stratton.

Rothbart, M. K., & Posner, M. I. (1985). Temperament and the development of self-regulation. In L. C. Hartlage & F. C. Telzrow (Eds.), *The neuropsychology of individual differences: A developmental perspective* (pp. 93-123). New York, NY: Plenum Press.

Rothbart, M. K., Taylor, S. B., & Tucker, D. M. (1989). Right-sided facial asymmetry in infant emotional expression. *Neuropsychologia, 27*, 675-687.

Sackeim, H. A., Greenberg, M. S., Weiman, A. L, Gur, R. C., Hungerbuhler, J. P., & Geschwind, N. (1982). Hemispheric asymmetry in the expression of positive and negative emotions. *Archives of Neurology, 39*, 210-218.

Sackeim, H. A., & Gur, R. C. (1978). Lateral asymmetry in intensity of emotional expression. *Neuropsychologia, 16*, 473-481.

Sackeim, H. A., & Gur, R. C. (1980). Asymmetry in facial expression. *Science, 209*, 834-836.

Safer, M. A. (1981). Sex and hemisphere differences in access to codes for processing emotional expressions and faces. *Journal of Experimental Psychology: General, 110*, 86-100.

Segalowitz, S. J. (1985). Hemispheric asymmetries in emotional faces. Paper presented at the *International Neuropsychology Society* meeting, February 6-8, San Diego.

Silberman & Weingartner (1986). Hemispheric lateralization of functions related to emotion. *Brain and Cognition, 5*, 322-353.

Sroufe, L. A. (1979). Socioemotional development. In J. Osofsky (ed.), *Handbook of infant development.* New York: John Wiley.

Strauss, E., & Moscovitch, M. (1981). Perceptual asymmetries in processing facial expression and facial identity. *Brain and Language, 13*, 308-322.

Terzian, H. (1964). Behavioral and EEG effects of intracarotid sodium amytal injection. *Acta Neurochirurgica, 12*, 230-239.

Trope, I, Rozin, P, Kemler Nelson, D., & Gur, R. C. (1992). Information processing in the separated hemispheres of callosotomy patients: Does the analytic-holistic dichotomy hold? *Brain and Cognition, 19*, 123-147.

Tucker, D. M. (1981). Lateral brain function, emotion, and conceptualization. *Psychological Review, 89*, 19-46.

Wirsen, A., af Klinteberg, B., Levander, S., & Schalling, D. (1990). Differences in asymmetric perception of facial expression in free-vision chimeric stimuli and reaction time. *Brain and Cognition, 12*, 229-239.

FOOTNOTES

[*] *Journal of Experimental Psychology: Human Perception and Performance*, in press.

[†] Also Wesleyan University.

[‡] Wesleyan University.

[1] Thus, these observations call into question the assumptions that AU6-7 activity is an unequivocal marker of spontaneous or genuine smiles, and/or that felt smiles are symmetrical in expressiveness. Eye crinkling apparently can also occur with socially-conditioned, elicited smiles, and these smiles do show perceived expressive asymmetries. This illustrates some of the difficulties inherent in relying solely on physical measures to assess the emotionality of expressions and the motivations behind them.

[2] The smiling infant in the figure is one of the two posers who showed AU6-7 activity around the eyes. The other four smiling infants showed none of the AU6-7 activity that is thought to reflect "felt" smiles even in infants (e.g. Fox & Davidson, 1988).

Haskins Laboratories Status Report on Speech Research
1993, SR-114, 159-166

On Determining the Basic Tempo of an Expressive Music Performance*

Bruno H. Repp

In an expressive music performance, the local tempo varies continuously and often asymmetrically around an implied (nominally constant) basic tempo. This preliminary study explored how pianists organize the expressive timing structure around an intended basic beat rate, how listeners judge the basic tempo of such a modulated performance, and what objectively measurable property of the performances the intended and/or judged tempi might correspond to. Two pianists played Robert Schumann's "Träumerei" three times at each of three intended tempi cued by a metronome. Tempo judgments (metronome settings) for the initial 8 bars of each performance were subsequently obtained from listeners who were pianists themselves. The judged tempi were generally slower than the intended tempi, which was attributed to a tendency of the performers to play slower than intended, especially at the faster tempi. The timing microstructure of each performance was quantified in terms of the frequency distribution of (raw or transformed) beat inter-onset intervals (IOIs). The judged tempi were generally close to the mean of this distribution (transformations had little effect on the mean tempo), which thus seems to be the parameter that best corresponds to the perceived beat rate of an expressively modulated performance, at least when there are no extreme *ritardandi*.

INTRODUCTION

The tempo at which a piece of music is to be performed is often indicated in the composer's score by a metronome (M.M.) number, such as "M.M.=60," meaning 60 beats per minute or a beat duration of 1 second. Such an instruction is easy to follow when the music in question has a steady rhythm; if necessary, the obedient performer can practice with the metronome ticking, aligning beats with ticks. Similarly, it is easy to determine the tempo of such a performance by finding the metronome rate that aligns itself with the beats, or by counting the number of beats in one minute of music.

When the beat is relatively slow and the music is highly expressive, however, the tempo is not so easily implemented or determined. Expression calls for considerable deviations from rigidity in timing, and these deviations are more often lengthenings than shortenings of beat durations, because *ritardandi* have the important function of marking structural boundaries at several levels (Repp, 1992; Todd, 1985). As a result, a count of the number of beats per minute may underestimate the tempo. Nor is it possible to align a metronome with the beats: Since an expressive performance calls for a continuous modulation of the tempo, it is virtually impossible to find a stretch of music during which the tempo is constant, and the occurrence of *ritardandi* destroys the synchrony between metronome and music. What, then, is the tempo of such a performance? And how does a performer implement an intended tempo *con espressione*?

It might be argued that such performances do not have a basic tempo. This objection can be dismissed, however, because even highly

---

This research was made possible through the generosity of Haskins Laboratories (Carol A. Fowler, president). Additional support came from NIH Grants RR05596 and MH51230. I am grateful to LPH for her patient participation in this study, and to Caroline Palmer, Henry Shaffer, and an anonymous reviewer for helpful comments on earlier versions of the manuscript.

expressive pieces are commonly preceded by metronome indications in the score. If composers and editors think such music has a tempo, there must be a principled (if intuitive) way of following their tempo prescriptions. Conversely, musicologists and music critics are often interested in how the tempo of a performance compares with the metronome number in the score. Thus, there should be a way of determining the underlying beat rate (M.M. number) of even a highly expressive performance. The problem of identifying the underlying or basic tempo of a performance is also relevant to methods of "quantization" in automatic rhythm detection and computer transcription of music (see, e.g., Desain & Honing, 1989, 1992), to theories of human rhythm perception (e.g., Jones & Boltz, 1989; Desain, 1992), and to performance modelling and synthesis (e.g., Todd, 1985).

That this is a nontrivial problem which apparently has not been addressed directly in the literature became evident to the author during a recent analysis of the expressive timing patterns in 28 performances of Robert Schumann's well-known piano piece, "Träumerei" (Repp, 1992). Most editions of the score contain either of two tempo prescriptions, one (M.M.=100) being attributed to the composer and the other (M.M.=80) to his wife, the pianist Clara Schumann. Questions of authenticity aside, both tempi seem unusually fast to contemporary ears (cf. Brendel, 1981, 1990; but see also Csipák & Kapp, 1981). Most, probably all, of the 28 performances examined by Repp (1992) were slower than M.M.=80. But what exactly *were* their tempi?

Informal clues to the intended tempi of two of these performances were available. In an article written at about the time his recording was made, Alfred Brendel (1981) mentions that his own preferred tempo for "Träumerei" is M.M.=69 (a statement repeated in Brendel, 1990). Another recording was by Fannie Davies, who had been a pupil of Clara Schumann. Although her performance was recorded much later in her career (in 1928), it was the fastest in the set, which suggested that she may have intended to adhere to her teacher's recommendation of M.M.=80. By a curious coincidence, the reciprocals of both of these informal tempo estimates happened to correspond to the 16th percentile of the total beat inter-onset interval (IOI) distribution of each pianist's performance. This percentile was unexpectedly low, however, suggesting that both informal tempo estimates may have been too high. (Fast tempi imply short IOIs.)

Nevertheless, these informal observations suggested a hypothesis: that the tempo of an expressive performance might be best characterized by an M.M. number corresponding to some fixed point along the IOI distribution, perhaps the median (50th percentile) or some point below it, or the mode (most frequent value). Alternatively, while the arithmetic mean of a skewed IOI distribution underestimates the basic tempo, the possibility remains that the mean of a *transformed* IOI distribution comes close to the "real" tempo. Reasonable choices of transformations are logarithms (the antilogarithm of whose mean is the geometric mean of the original IOI distribution) and reciprocal values, which (when IOIs are expressed in fractions of a second) represent estimates of *local tempo*. A logarithmic transformation could be justified on the basis that it takes into account Weber's law, which holds approximately for duration discrimination above 300 ms (e.g., Drake & Botte, 1993). In fact, Wagner (1974) used the geometric mean to estimate tempo, based on this consideration. The reciprocal transformation seems reasonable because it represents tempo directly. Both transformations have the effect of reducing the asymmetry of the IOI distribution. As to the location of a given tempo estimate on the cumulative IOI distribution, it should be noted that it depends only on the order of IOI values and therefore is unaffected by any monotonic transformation (though the reciprocal transformation reverses the order of values).

The following exploratory experiment investigated the relationships among (1) the tempo intended by the performer, (2) the actual timing microstructure of the performance, and (3) the tempo judged by musically trained listeners. The specific question of interest was whether the underlying tempo of a performance can be expressed in terms of an invariant statistical parameter of the (original or transformed) IOI distribution.

The music was again Schumann's "Träumerei." The study focused on the initial section of the music, which is shown in Figure 1. This section begins with the upbeat in "bar 0" and ends with the chord on the third beat of bar 8, a total of 32 quarter-note beats. There were three reasons for restricting attention to this section: (1) The performers, who intended to follow a tempo cued by a metronome, naturally remembered the tempo best at the beginning of the performance and were expected to be most accurate there. In fact, there is evidence from a more detailed analysis of these and other performances of "Träumerei" that the tempo usually slows down later in the piece, presumably for expressive reasons (Repp, 1992, 1994). (2) Performance excerpts of the same length were used in the tempo judgment task described below, to limit the duration of the test. (3) The initial section does not contain any extreme *ritardandi*, such as typically occur at the end of the middle section (bar 16) and at the end of the piece (bars 23-24). It seems likely that players and listeners would not include such extreme tempo changes in their mental estimates of the basic tempo. It seemed that the first 8 bars contain sufficient local tempo variation to address the question posed here in a meaningful, if preliminary, way. (For plots of the "timing profiles" of the performances analyzed here, see Repp, 1994.)

PERFORMANCE ANALYSIS

Methods

Pianists. Two pianists participated. One was a professional musician (LPH) in her mid-thirties; the other was a serious amateur (BHR, the author) in his late forties. Both were thoroughly familiar with Schumann's "Träumerei" and had played it many times previously.

Equipment. The instrument was a Roland RD250S digital piano connected to a microcomputer that registered performance data in MIDI format (onset time, offset time, and key velocity), with a temporal resolution of 5 ms. "Piano 1" sound was used, and a simple on/off sustain pedal switch (DP-2) was taped to the floor. The sound output was monitored over earphones by the performer. A brand new Franz LM-FB-4 electric metronome stood nearby on a table.

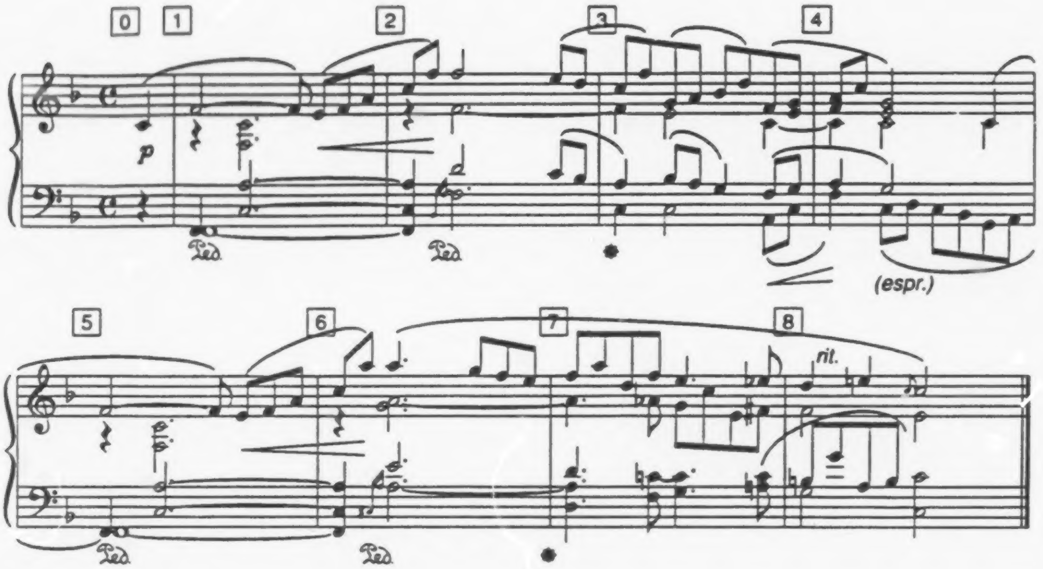


Figure 1. The initial 8 bars of Schumann's "Träumerei," with the final chord extended through the second half of bar 8.

Procedure. Each pianist performed the complete piece 9 times from the score, three times at each of three intended tempi. At the beginning of the recording session, she/he warmed up on the keyboard and then played the piece once at her/his preferred tempo while being recorded. The beginning of this MIDI recording was subsequently played back, and the pianist set the metronome to the beat frequency that best corresponded to the tempo of the performance (as in the tempo judgment task described below). The settings chosen by LPH and BHR were M.M.=63 and 66 ("medium tempo"), respectively. The recording of this initial performance was discarded.

The desired tempo for each subsequent performance was indicated by the metronome, which was left running at the desired speed for a while and was turned off just before each performance started. "Slow" and "fast" metronome settings were chosen by the author to surround the medium tempi: M.M.=54 and 72, respectively, for LPH, and M.M.=56 and 76, respectively, for BHR. These tempi were intended to be within the range of aesthetic acceptability for "Träumerei" (cf. Repp, 1992) and thus did not force the pianists to play in an unnatural manner. LPH played in the order slow-fast-medium (repeated twice), whereas BHR played in the order medium-slow-fast (repeated twice). The performances were free of hesitations and major technical accidents, and were judged by the author to be appropriately expressive renditions of the score. (See Repp, 1994, for a more detailed analysis of their expressive microstructure.)

Analysis. To determine beat (quarter-note) onset times, the tone with the highest pitch in any cluster of nominally simultaneous tones falling on a beat was picked. Inter-onset intervals (in milliseconds) were subsequently computed from this reduced "MIDI score." Missing beat onsets (of which there were 4; see Figure 1) were interpolated by subdividing longer IOIs into smaller intervals of equal duration. Average performances for each intended tempo were obtained by linearly averaging the corresponding IOIs of each pianist's three individual performances. The raw individual and average IOIs were also transformed into logarithms and into local tempo estimates (beats per minute, M.M.=60000/IOI). Means were calculated and, for raw and logarithmic IOIs, transformed into tempo estimates (M.M.=60000/mean and M.M.=60000/e^{mean}, respectively).

Results

Figure 2 shows the local tempo distributions of the two pianists' average performances at each of the three intended tempi. Local tempi varied by as much as 30 beats per minute. As expected, the distributions shifted towards faster tempi as the intended tempo increased. The solid vertical line indicates the intended tempo. Its location does not generally coincide with the mode of the distribution.

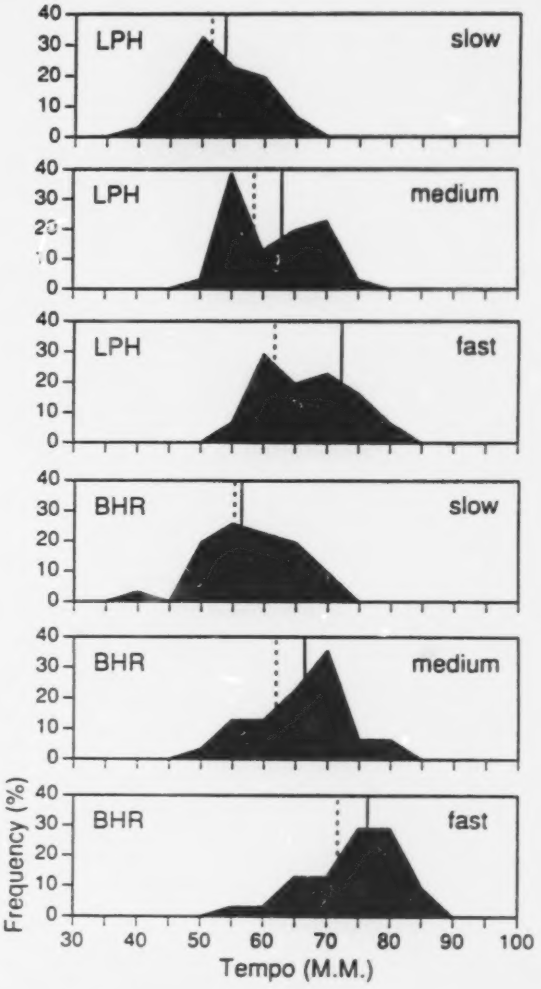


Figure 2. Frequency distributions of local tempi at each of three intended tempi for the two pianists, based on IOIs averaged over the three performances at each intended tempo. The bin width is 5, and frequency values are plotted at the upper limits of bins. Solid vertical lines indicate intended tempi, dotted lines judged tempi.

Figure 3a shows the percentiles of the cumulative local tempo distributions of the 18 individual performances that correspond to the intended tempi. It is evident that there is no fixed point along these distributions that characterizes the intended tempo. For each pianist the percentile increases as the intended tempo increases. Also, there is a large difference between the two pianists, with LPH showing higher percentiles than BHR, and there is considerable variability within each tempo category.

Figure 3b compares the intended tempo with the mean of the local tempo distribution for each of the 18 performances. It is evident that the mean consistently underestimates the intended tempo, more so for LPH than for BHR, and more so at fast than at slow intended tempi. (Analogous plots using tempo estimates based on the arithmetic or geometric mean of the original IOI distribution show similar but slightly larger differences.) The figure also suggests that LPH played relatively slower than BHR, even when the differences in intended tempi are taken into account.

The intended tempi thus do not correspond to any invariant parameter of the IOI distribution. However, it is possible that the two pianists did not implement the intended tempi accurately. In particular, they may have played slower than intended, and more so at fast than at slow tempi. Perceptual estimates of the tempi of these performances may shed light on this issue.

PERCEPTUAL JUDGMENTS

Methods

Listeners. The listeners were nine skilled pianists, most of them graduate students of piano performance at the Yale School of Music, who were paid for their services. They were tested individually in a 1-hour session that began with a different task using the same materials (Repp, 1994).

Procedure. The initial 8-bar sections of the 18 performances were excerpted and stored in separate MIDI sequence files. The final chord was extended to provide a pleasing conclusion (cf. Figure 1). The author, who conducted the experimental session, called up these files in a different sequence for each listener, according to a counterbalanced schedule. Each sequence was constructed so that performances by the two pianists alternated, performances in the same tempo category did not follow each other, and each block of 6 included one performance of each pianist in each intended tempo category. The listener sat in front of a computer keyboard, wore earphones connected to the digital piano, and manipulated the metronome on the table next to the keyboard.

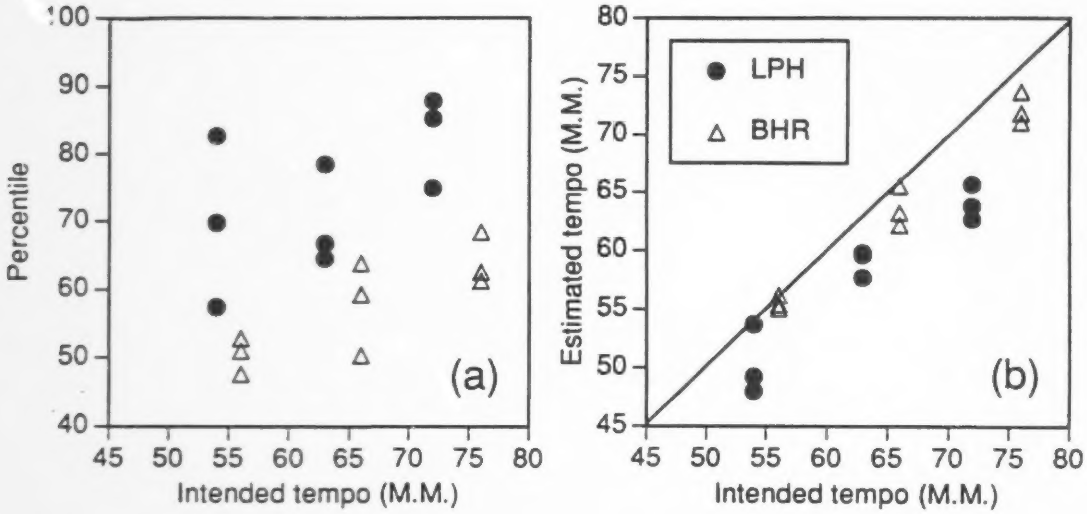


Figure 3(a). Percentiles of the cumulative local tempo distributions corresponding to intended tempi, for all individual performances. *3(b).* The relationship between intended tempo and the mean of the local tempo distribution. The diagonal line indicates equality.

He/she was shown how to start, stop, and restart MIDI playback by pressing certain keys. The task was to find the beat frequency that best approximated the tempo of each performance by adjusting the metronome and reporting the M.M. number, which was recorded by the author. The strategy for accomplishing the task was left up to the listener. He/she could take as much time as necessary, start and stop MIDI playback at will, use the metronome in sound (click and flash) or silent (flash only) mode, and adjust it while the music was on or off. Most listeners repeatedly alternated between listening and adjustment periods and took about 1 minute per judgment. The resolution of the metronome in the region of interest was 2 steps (beats per minute) up to M.M.=60, 3 steps up to M.M.=72, and 4 steps above that.

Results

The tempo judgments were averaged over listeners. Their average standard deviation was 4 metronome steps, so that the average standard error was about 1.3 steps. The judged tempi are shown as a function of intended tempi in Figure 4. It is evident that most performances were judged to have slower tempi than the pianists had intended, particularly when the intended tempo was fast. In fact, the pattern of these data is quite similar to that in Figure 3b. Although, in principle, the discrepancies between intended and judged tempo could represent systematic errors of judgment (i.e., tempo underestimation), it seems more plausible that they reflect performance deviations. After all, there were nine judges but only one pianist per performance.

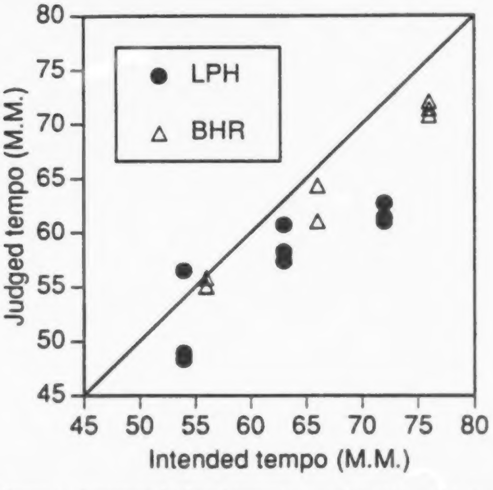


Figure 4. Average judged tempo as a function of intended tempo for individual performances. The diagonal line indicates equality.

Assuming, then, that the listeners' average judgments represent reasonably accurate estimates of the basic tempi of these performances, we may ask now whether *they* correspond to a particular parameter of the tempo distributions. Figure 5 shows the percentiles of the tempo distributions that correspond to the judged tempi. It can be seen that there is still no constancy, especially not for pianist LPH. A glance back at Figure 2 also shows that the judged tempi (vertical dotted lines) do not generally coincide with the modes of the tempo distributions.

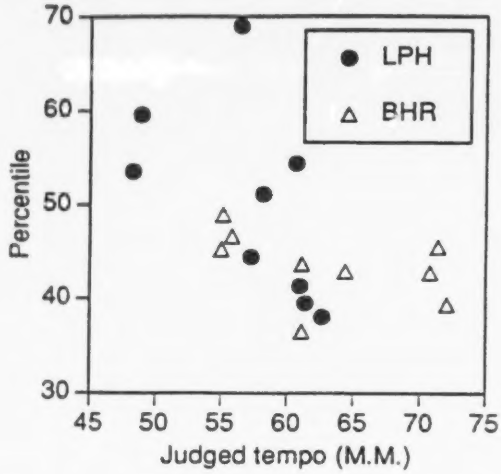


Figure 5. Percentiles of the cumulative local tempo distributions corresponding to judged tempi, for all individual performances.

Figure 6 compares the judged tempi with the mean tempi, calculated either from the raw IOIs (Figure 6a) or from the local tempo estimates (Figure 6b); results based on the logarithmic IOIs are extremely similar to those in Figure 6b. Here there is a very satisfactory match. There were only small differences among the three types of objective tempo estimates, due to the absence of extreme asymmetries in the IOI distributions. The estimates in Figure 6a are slightly lower than those in Figure 6b, but the match with judged tempi is comparably good. Thus, these data do not favor a particular transformation of the IOI values. Rather, they suggest that any estimate of average tempo is a reasonable approximation of the basic (perceived) tempo of an expressively modulated performance. One may anticipate, however, that the mean of the local tempo distribution (Figure 6b) or the reciprocal geometric mean IOI will be preferable for more strongly skewed IOI distributions.

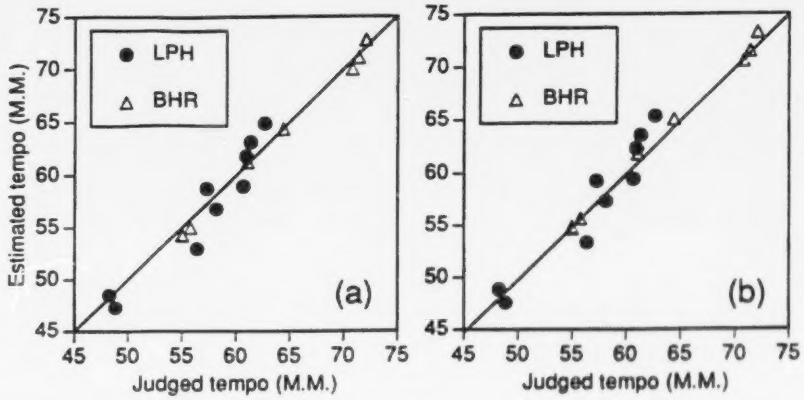


Figure 6(a). **Judged tempo versus mean tempo based on raw IOIs.** *6(b).* **Judged tempo versus mean tempo based on local tempo estimates. The diagonal line indicates equality.**

DISCUSSION

This study addressed a question that apparently has not been asked previously in the psychological or musicological literature, though it seems related to the problem of "quantization" in computer music applications (see Desain & Honing, 1989, 1992). Quantization algorithms attempt to recover a rigid metrical structure from an expressively modulated performance. Simple algorithms assume a constant tempo (a metric grid) and consequently make many errors when the timing modulations are large and/or asymmetric, as they often are. More sophisticated algorithms attempt to track tempo changes. In doing so, they negate (or at least do not address) the idea of a single basic tempo. Thus the goal of quantization is somewhat different from the question pursued in the present study.

In studies of music performance, measured IOIs have often been plotted as percentage deviations from a horizontal baseline (see, e.g., Gabrielsson, 1987; Palmer, 1989). This tradition goes back to Seashore's (1938, p. 9) famous dictum that musical expression is "deviation from the regular." "The regular," in the case of timing, is a mechanically exact rendition of the underlying beat of the expressively modulated performance. In the performance studies referred to, the baseline was placed at the mean IOI. Somewhat unexpectedly, the present data seem to vindicate this procedure. However, it must be remembered that the excerpt investigated here did not contain extreme *ritardandi,* which would have pulled down the estimate of average tempo. Therefore, the author still prefers to avoid a baseline and to plot original IOIs on a log scale, which conveys relative as well as absolute magnitudes (see Repp, 1992, 1994).

The absence of extreme asymmetries in the present IOI distributions raises the question of whether the present findings would generalize to musical excerpts containing severe *ritardandi.* On one hand, there surely would be a larger difference between arithmetic, geometric, and local tempo means, probably favoring the latter. On the other hand, it seems implausible that a listener would include such extreme slow-downs in his or her estimate of basic tempo. Presumably there is a limit to what a listener is willing to accept as being played at the same basic tempo; beyond this limit, the tempo is simply perceived as changing or different. A more precise determination of this limit must await further research.

Some ambiguity remains in the present data, for as long as the basic tempo is not known, it is impossible to separate performance inaccuracy from systematic biases in tempo judgment. Although it seems rather clear that the present performers played slower than intended, there is no guarantee that the listeners judged the tempo accurately. For example, it is possible that the pianists, rather than playing too slow across the board, drifted towards their preferred (medium) tempo. If so, then the listeners must have consistently underestimated the basic tempo. This in turn might account for the unexpectedly close match of the judged tempi with the tempo estimates derived from the arithmetic mean IOI (Figure 6a). A tendency of musicians to underestimate tempi has been reported previously by Madsen (1979). Clearly, this issue requires further research.

Repp (1992: Table III) based his tentative tempo estimates for 28 performances of "Träumerei" on the first quartiles (25th percentiles) of the distributions of eighth-note (half-beat) IOIs for the whole music. (The Brendel and Davies data, originally analyzed in this format, had suggested percentiles near the first quartile.) It seems now that this measure probably overestimated the tempo, though the inclusion of the later sections of the piece, with their somewhat slower tempo and large *ritardandi,* may have reduced the error. Revised estimates representing the mean local tempo during the first rendition of the initial 8-bar section, based on quarter-note (beat) IOIs, are indeed much slower than those reported in Repp (1992). Contrary to his own stated preference of M.M.=69, Brendel's tempo is only M.M.=58, and Davies's tempo, at M.M.=71 still the fastest in the set, falls short of Clara Schumann's recommended M.M.=80. Thus these initial clues towards defining the basic tempo appear to have been misleading: According to the present data, the basic tempo does not correspond to a particular percentile or the mode of the IOI distribution but rather to the mean of the (transformed) distribution. However, this conclusion will have to be tested further with more extensive sets of data.

REFERENCES

Brendel, A. (1981). Der Interpret muß erwachsen sein: Zu Schumanns "Kinderszenen." *Musica, 35,* 429-433.

Brendel, A. (1990). Testing the grown-up player: Schumann's 'Kinderszenen.' In A. Brendel, *Music sounded out* (pp. 142-153). New York: Noonday Press. [Revised and translated version of Brendel, 1981.]

Csipák, K., & Kapp, R. (1981). Träumerei. *Musica, 35,* 438-443.

Desain, P. (1992). A (de)composable theory of rhythm perception. *Music Perception, 9,* 439-454.

Desain, P., & Honing, H. (1989). The quantization of musical time: a connectionist approach. *Computer Music Journal, 13,* 56-66.

Desain, P., & Honing, H. (1992). *Music, mind and machine.* Amsterdam: Thesis Publishers.

Drake, C., & Botte, M.-C. (1993). Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. *Perception & Psychophysics, 54,* 277-286.

Gabrielsson, A. (1987). Once again: The theme from Mozart's Piano Sonata in A major (K.331): A comparison of five performances. In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 81-104). Stockholm: Royal Swedish Academy of Music (Publication No. 55).

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review, 96,* 459-491.

Madsen, C. K. (1979). Modulated beat discrimination among musicians and nonmusicians. *Journal of Research in Music Education, 27,* 57-67.

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 331-346.

Repp, B. H. (1992). Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's "Träumerei." *Journal of the Acoustical Society of America, 92,* 2546-2568.

Repp, B. H. (1994). Relational invariance of expressive microstructure across changes in global tempo: An exploratory study using Schumann's "Träumerei." *Psychological Research,* in press.

Seashore, C. E. (1938). *Psychology of music.* New York: McGraw-Hill. (Reprinted in 1967 by Dover Publications.)

Todd, N. (1985). A model of expressive timing in tonal music. *Music Perception, 3,* 33-58.

Wagner, C. (1974). Experimentelle Untersuchungen über das Tempo. *Österreichische Musikzeitschrift, 29,* 589-604.

FOOTNOTE

[*] *Psychology of Music,* in press.

Haskins Laboratories Status Report on Speech Research
1993, SR-114, 167-178

Musical Motion: Some Historical and Contemporary Perspectives*

Bruno H. Repp

The idea that music "moves" has a long and varied history. Some aspects of this notion are metaphorical (e.g., the "motions" between pitches, harmonies, or keys), whereas others are more literal. The latter derive from the performer's actions that bring the music to life. This gestural information is encoded in the expressive microstructure of the performance at several hierarchically nested levels. Some older demonstrations in support of this proposition have used the technique of "accompanying movements", devised and elaborated by authors such as Eduard Sievers, Gustav Becking, and Alexander Truslit. Contemporary approaches, most notably those of Manfred Clynes and Neil Todd, focus instead on performance analysis and synthesis. Todd has provided evidence that tempo modulations in expert performances obey a constraint of linear changes in velocity, suggesting that music is "set into motion" by some kind of force or impulse function. Clynes has proposed (following Becking) that the parameters of these underlying functions distinguish different composers. The notion of *spatio-temporal coupling*, illustrated by Paolo Viviani's work on drawing movements, suggests a theoretical basis for the recovery of spatial movement from temporal information. Physical laws of motion thus impose constraints on performance microstructure, constraints that are also reflected in listeners' perception and aesthetic judgments.

INTRODUCTION

Music is made by moving hands, fingers, or extensions thereof over an instrument, and the dynamic time course of these movements is reflected to some extent in the resulting stream of sounds. Conversely, people listening to music frequently perform coordinated movements that range from foot tapping to elaborate dance. Although these movements on the listener's side are not the same as those of the performer, they are certainly not unrelated. At the very least, they share a rhythmic framework that gets transmitted from player to listener via the sound structure.

In many cultures this close connection of music and movement is so obvious as to hardly deserve comment. In Europe, however, the remarkable development of musical notation and of complex compositional techniques over the last few centuries has led to a focus on the structural rather than the kinematic properties of music, at least of so-called serious music. At the same time, as this music was performed mainly in church or concert halls, a social restriction against overt movement in listeners has long been in effect. As a result of these practices, the close connection of music and motion has receded from people's consciousness, and 20th century aesthetic and technological developments have occasionally even severed that connection, with only few taking notice. Therefore, there is a need today to reassess the concept of musical motion and its role in performance and music appreciation.

My purpose in this paper is not to review philosophical or musicological treatments of this topic; suffice it to mention the important discussions by Langer (1953), Zuckerkandl (1973), and Sessions (1950), among many others. Rather, I will focus on the limited and far-between attempts to provide empirical demonstrations of the kinematic correlates of Western art music. Also, I will not dwell on the more abstract and metaphoric notions of melodic and harmonic motion common among musicologists, which concern the transitions from one pitch, or one harmony, or one tonality to another--movements that can be seen, as it were, by moving one's eyes over the printed score. I am concerned primarily with *rhythmic* motion, which presupposes a *performance*, a human realization of the music as structured sound, whether actual or imagined. The question I am pursuing, then, is: What is the nature of the rhythmic motion information in music, and how can its kinematic implications be demonstrated?

In this presentation I intend to review briefly the pioneering work of three largely forgotten individuals who were active in Germany during the early decades of this century. In doing so, I hope to inform or remind you of their theoretical accomplishments, however limited their empirical contribution may seem from our modern scientific perspective. Then I will turn to sampling the work of two contemporary researchers who—knowingly in one case, unwittingly in the other—have elaborated upon and increased the precision of the German pioneers' ideas, so that they can now be subjected to rigorous tests. I will conclude with a very brief foray into the motor control literature, again focusing on a single researcher whose work seems to be particularly pertinent to the kinds of motion that music engenders. Because of time constraints I will not be able to do justice to the related work of many others, for example Johan Sundberg and Alf Gabrielsson; to them I apologize, but you can hear about their latest work first-hand at this conference.

Three German pioneers: Sievers, Becking, and Truslit

Whereas no one doubts that there is visual information for motion, the concept of auditory motion information is less widely accepted, especially since it involves an essentially stationary sound source—the musical instrument being played on. One reason for this scepticism may be that visual motion information is generally continuous in time, whereas auditory motion information, especially that in music, is often carried by discrete events (i.e., tone onsets) that only *sample* the time course of the underlying movement. The principal technique for demonstrating that music does convey movement information is the reconstitution of the analogous spatial movement by a human listener. The listener's body thus acts as a *transducer*, a kind of filter for the often impulse-like coding of musical movement.

The first modern attempt to use such a technique in a systematic fashion must be credited to the German philologist Eduard Sievers, who applied it not to music but to literary works. Sievers called his method *Schallanalyse* ("sound analysis"), though it was not concerned with sound as such but rather with body posture and movement as a way of reconstructing and analyzing the expressive sound shape of printed language, mostly poetry. He never published a complete account of his very complex methods. Sievers (1924) provides an overview; for a more recent critical evaluation, see Ungeheuer (1964).

Sievers's initial impetus came from observations of a teacher of singing, Joseph Rutz, published by his son Otmar Rutz (1911, 1922), about connections between body posture and voice quality. Certain body postures were said to inhibit vocal production, whereas others facilitated it and gave it a free, uninhibited quality. Sievers initially focused on these static body postures which he symbolized by means of "optic signals" in the form of geometric shapes that were meant to cue different body postures in a speaker reciting a text. Subsequently he elaborated this method into a system of dynamic movements, to be carried out with a baton, with the index finger, or even with both arms while speaking. The crucial criterion was the achievement of "free and uninhibited articulation", and the goal of the analytic method was to find the accompanying movements that interfered least with (or facilitated most) the recitation of the text. The metric, prosodic, and semantic characteristics of the text naturally varied across authors and their individual works, and so did the accompanying movements considered optimal by Sievers. The movements were rhythmically coordinated with the speech and had a cyclic or looping character. However, they could vary in a large number of features, such as the relative smoothness of turns, the tilt of the main axis, rising versus falling direction, etc.

Sievers distinguished two classes of curves, which are illustrated in Figure 1: general curves or "Becking curves", and specific or filler curves (*Taktfüllcurven*). The former, which were suggested to him by Gustav Becking (see below), come in three types that in fact exhaust the possibilities for a cyclic movement with two turning points: pointed-round, round-round, and pointed-pointed. Any individual speaker/writer was said to be characterized by one and only one of these types, if not as an obligatory then at least as a preferred mode of dynamic expression, and hence by a corresponding "voice type". However, many variations are possible within each type. The "special curves", of which there is a bewildering variety, reflect the specific metric and sonic properties of a spoken text (or of music, as the case may be). It was these special curves and their many variations that Sievers devoted most of his efforts to.

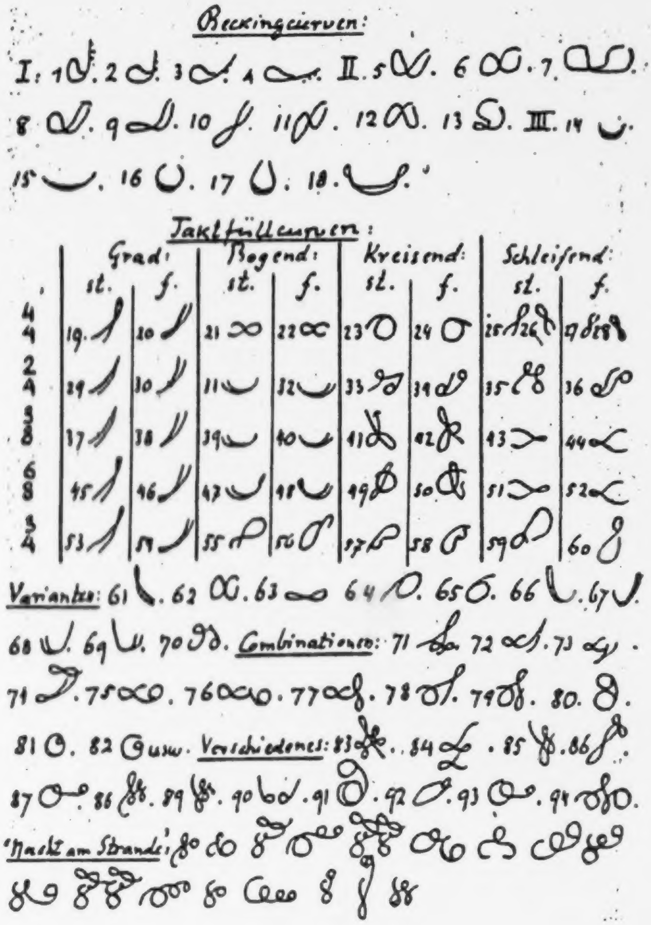


Figure 1. **Examples of movement curves used by Eduard Sievers. Top: general curves. Center: special curves (straight, curved, circular, looping). Bottom: variations, combinations, miscellaneous, and a kinematic interpretation of a text, "Markt am Strande". (Reproduced from Sievers, 1924: 73.)**

Sievers was the only recognized master of the technique he had developed. He claimed to be in possession of an extraordinary "motoric sensibility" that, combined with many years of self-training and observation, enabled him to find the accompanying movements for the most subtle variations in the sound shape of spoken texts. Although his dedication and expertise were never in doubt, the extreme subjectivity of his method obviously reduced its respectability as a scientific procedure. Nevertheless, the basic idea underlying it remains of value: He showed that rhythmic sound patterns have a dynamic time course that can be translated into accompanying body movements. Only the rules governing this translation remained somewhat obscure.

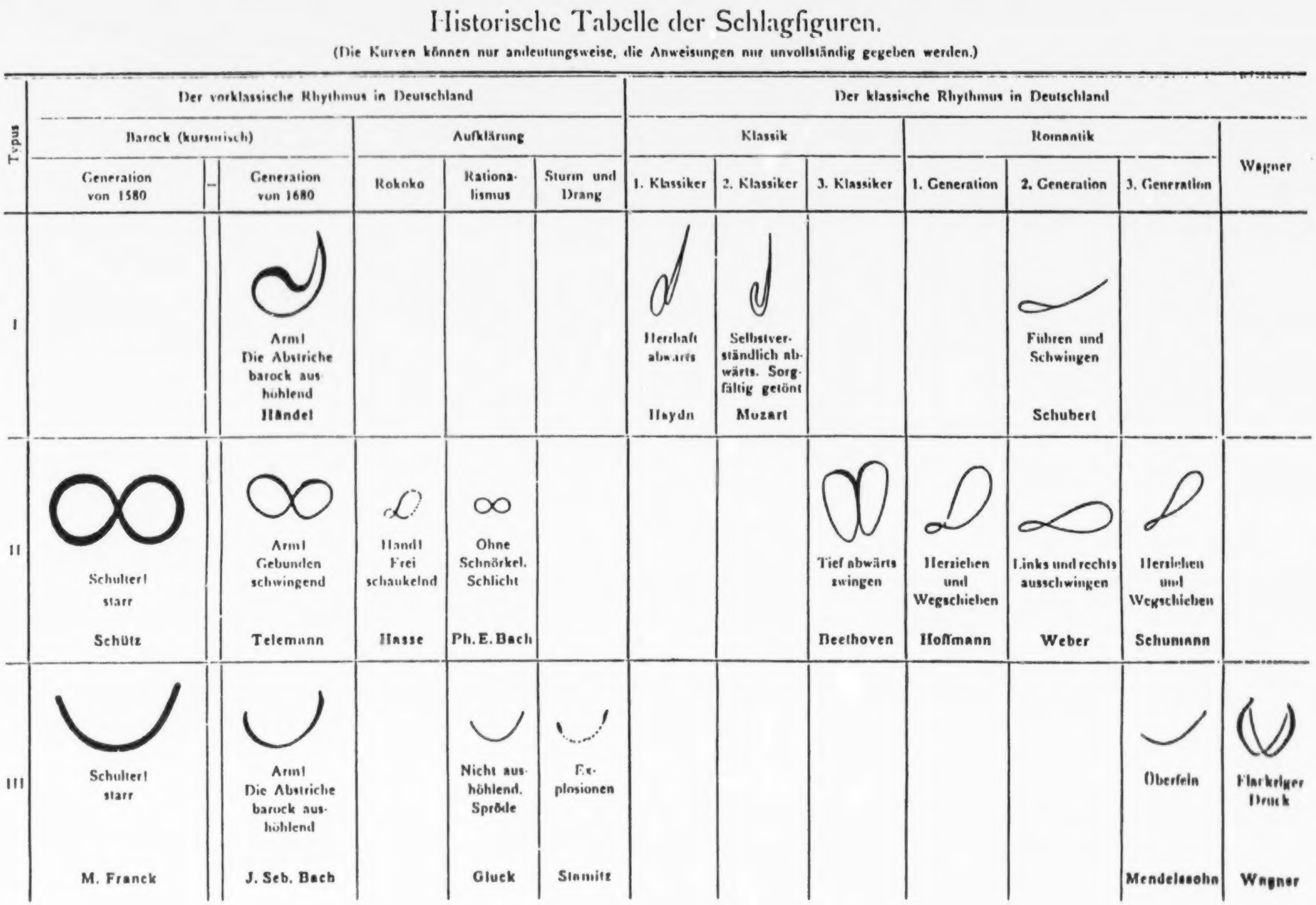
Sievers benefited from his interaction with Gustav Becking, a young musicologist who developed his own ideas in a monograph entitled *Der musikalische Rhythmus als Erkenntnisquelle* (Musical Rhythm as a Source of Insight) that appeared in 1928. Becking's pivotal assumption is that there is a *dynamic rhythmic flow* below the musical surface. This flow, a continuous up-down motion, connects points of metric gravity that vary in relative weight. Becking's important and original claim is that the distribution of these weights varies from composer to composer. The analytic technique for determining these weights is Sievers's method of accompanying movements, carried out with a light baton. A downbeat always accompanies the heaviest metric accent; then an upward movement follows which leads into the next downbeat. The *dynamic shape* of this movement cycle is of interest. For example, the strongest pressure in the downbeat is never at the beginning but at varying delays; the movement may be deep and vertical or shallow and more nearly horizontal; and the connection of down and up movements may be smooth or abrupt.

Becking's primary interest was not in the differentiation and proliferation of movement curves for individual works of art but in the personal constants of individual composers—in other words, invariance rather than variability. He says that the personal curves reflect a composer's individual "management of gravity". Gravity being a physical given, different composers' solutions reflect different philosophical attitudes towards physical reality—as something to be overcome, to adapt to, or to be denied, as the case may be. Becking's ultimate goal thus is a typology of personal constants linked to a typology of *Weltanschauungen* (world views)—a philosophical undertaking in which he was preceded by Nohl (1920), among others.

As already mentioned in connection with Sievers's "Becking curves", Becking distinguishes three types of "personal curves", examples of which are illustrated in Figure 2: *Type I* has a sharp, pointed onset of the downbeat, which is straight and usually vertical but nevertheless actively guided rather than passively falling. At the bottom, there is a narrow but round loop ending in a small downward movement (a secondary accent between downbeats) before leading vertically upward, resulting in a figure somewhat resembling a golf club. This pattern, with its strong differentiation of rhythmic accents but nevertheless organic dynamic shape is attributed to the "Mozart family", which also includes Handel, Haydn, Schubert, Bruckner, and most Italian composers. These composers are said to be monists (in that they largely obey the physical force of gravity) as well as idealists, because they actively impose a personal dynamic shape. *Type II* has a round, curving, inward-going (towards the body) onset of the downbeat and a similarly round, outward-going turn upwards, leading to a figure resembling a horizontal or tilted figure "8". Differences in accentuation among metric subdivisions tend to be reduced here. Composers characterized by this personal curve form the "Beethoven family", which also includes Weber, Schumann, Brahms, Richard Strauss, and many other German masters. According to Becking, they aim to overcome gravity and force it into a winding path. Thus they are dualists (in that they oppose the material force with their own spiritual force or will) as well as idealists (in that they impose an organic dynamic shape on the raw pulse of the music). Finally, *Type III* is characterized by a pointed downbeat as well as a pointed return, resulting in a semicircular, pendulum-like curving motion from right to left and back. Consequently, the main accents on the downbeats and the secondary accents in between tend to be equally strong and form a rigid rhythmic framework. This pattern Becking ascribes to the "Bach family", including Mendelssohn, Chopin, Wagner, Mahler, and most French composers. These composers are said to be naturalists because they follow the force of gravity without opposing it or necessarily imposing a personal pulse on it. Yet there are numerous personal variants of the trajectory between the two rigid endpoints, resulting in more or less idealistic curves (e.g., Wagner). Nevertheless, all these composers accept the objective, even pulse and hence are only minor idealists, with Bach being the least idealistic and most objective of all.

Becking's method of determining the personal curve of a composer was highly subjective. It required a thorough acquaintance with a composer's oeuvre as well as, presumably, with performances by great interpreters and biographical details that help elucidate the artist's personality. The personal curve is *not* derivable from the score, nor is its subjective fit to a particular piece of music necessarily perfect, especially if that music is an early or otherwise atypical creation. Rather, knowledge of the personal curve, verified on the composer's most characteristic works, enables the scholar or performer to imbue even the less characteristic works with the composer's identity. Clearly, this method is somewhat circular and not at all scientifically rigorous. However, Becking's extraordinary perspicacity, his well-chosen musical examples, and his eloquent verbal characterizations make his book a unique and fascinating document.

The third important person among the German pioneers is the one least known today—a man named Alexander Truslit, whose book, *Gestaltung und Bewegung in der Musik* (Shaping and Motion in Music), appeared in 1938. Truslit's orientation is much closer to the natural sciences than that of his predecessors and in some ways presages James Gibson's (1979) writings on ecological perception and action. Unlike Becking, who believed that composers' personal dynamics take place largely beneath the musical surface (i.e., in the listener's musical imagination), Truslit focuses on the information in the sound pattern.



Historische Tabelle der Schlagfiguren.

(Die Kurven können nur andeutungsweise, die Anweisungen nur unvollständig gegeben werden.)

Typus	Der vorklassische Rhythmus in Deutschland						Der klassische Rhythmus in Deutschland						
	Barock (kursorisch)			Aufklärung			Klassik			Romantik			Wagner
	Generation von 1580	–	Generation von 1680	Rokoko	Rationalismus	Sturm und Drang	1. Klassiker	2. Klassiker	3. Klassiker	1. Generation	2. Generation	3. Generation	
I			Arm! Die Abstriche barock aushöhlend **Händel**				Herzhaft abwärts **Haydn**	Selbstverständlich abwärts. Sorgfältig getönt **Mozart**			Führen und Schwingen **Schubert**		
II	Schulter! starr **Schütz**		Arm! Gebunden schwingend **Telemann**	Hand! Frei schaukelnd **Hasse**	Ohne Schnörkel. Schlicht **Ph. E. Bach**				Tief abwärts zwingen **Beethoven**	Herziehen und Wegschieben **Hoffmann**	Links und rechts ausschwingen **Weber**	Herziehen und Wegschieben **Schumann**	
III	Schulter! starr **M. Franck**		Arm! Die Abstriche barock aushöhlend **J. Seb. Bach**		Nicht aushöhlend. Spröde **Gluck**	Explosionen **Stamitz**						Überfein **Mendelssohn**	Flackriger Druck **Wagner**

Figure 2. **Becking's historical table of conducting curves for selected German composers. (Reproduced from the end plate in Becking, 1928.)**

He contends that the musical dynamics and agogics (i.e., timing variations) convey movement information directly to the sensitive listener, who can then instantiate these movements by acting them out, if necessary. The goal of music performance is to arrange the musical surface in accord with the appropriate underlying movement.

Like Becking, Truslit distinguishes three basic types of movement curves: "open", "closed", and "winding" (*gewunden*). In Figure 3, they (b-d) are contrasted with an unnatural linear motion path (a). Superficially, they resemble Becking's three types; in particular, the winding curve seems very much like Becking's Type II, and the open curve resembles Becking's Type III. However, these similarities are more apparent than real. Truslit's curves are not conducting movements; they are to be carried out slowly and with outstretched parallel arms, so that the whole upper body is involved. Their height in space tends to follow the pitch contour of the melody; thus they often start at the bottom and move upwards rather than beginning with a "downbeat". They are a means of portraying the melodic dynamics in space, with the speed of movement and the consequent relative tension being governed mainly by the curvature of the motion path. That is, a relative slowing down and increased tension in the music is portrayed by a tight loop, whereas faster, more relaxed stretches correspond to relatively straight movements. The varied melodic structure of a composition elicits complex paths of various combinations of clockwise and counterclockwise turns, interpolated loops, etc. Even the type of movement may change within a composition. Figure 4 illustrates the combination of closed and winding movements that Truslit found most appropriate for the initial section of Brahms' Rhapsody in G minor.

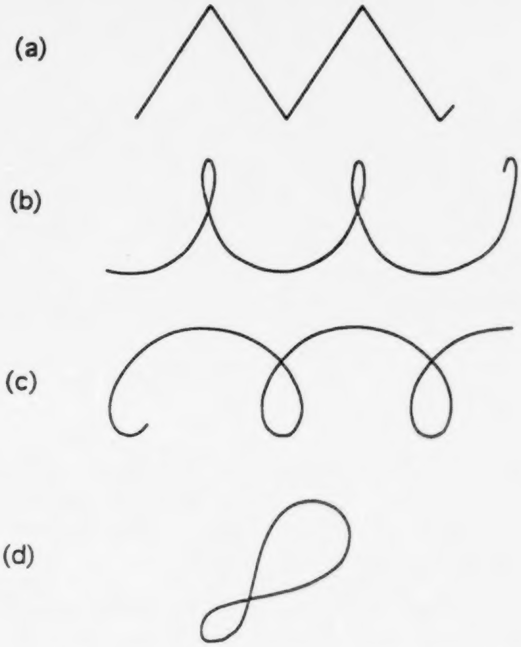


Figure 3. Truslit's movement types: (a) straight (mechanical); (b) open; (c) closed; (d) winding. (After Truslit, 1938: Plate 2; reproduced from Repp, 1993: 54, with permission of the publisher.)

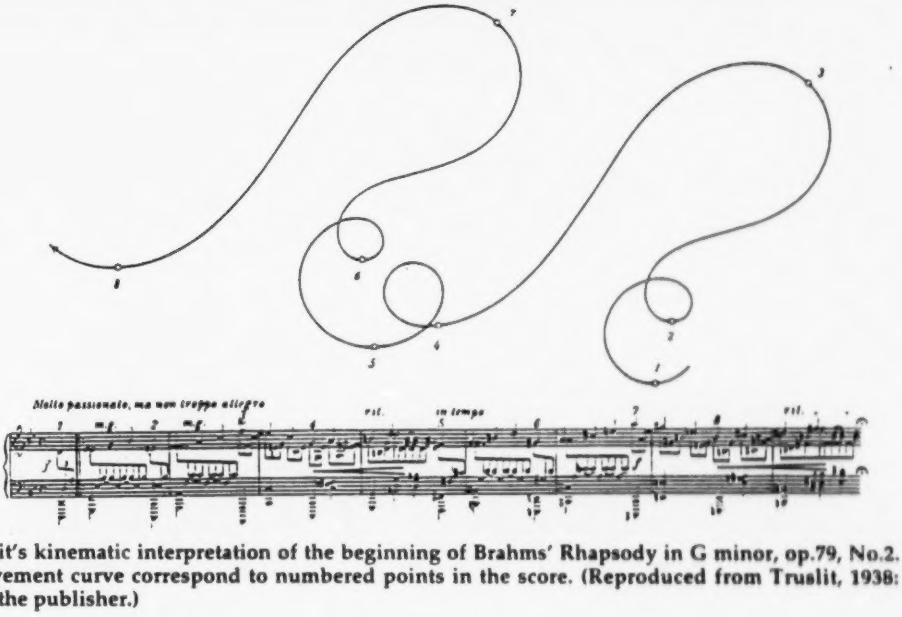


Figure 4. Truslit's kinematic interpretation of the beginning of Brahms' Rhapsody in G minor, op.79, No.2. Numbers along the movement curve correspond to numbered points in the score. (Reproduced from Truslit, 1938: 144, with permission of the publisher.)

Truslit's curves are not at all "personal" and composer-specific; rather, they are work-specific. In that respect, he is somewhat closer to Sievers than to Becking. He explicitly assigns a secondary and subordinate role to rhythmic patterns: They should not be too pronounced, so as not to disrupt the smooth flow of the melody. Rhythmic patterns affect the limbs, he says (which is consistent with Becking's use of the hand to conduct), whereas the more global melodic patterns affect the large muscles of the back and hence the whole body. Thus Truslit's curves often extend over a number of measures, with the more detailed rhythmic structure being marked by small local loops, if at all. Not surprisingly, Truslit seems to be most interested in music that exhibits a pronounced gestural character; many of his musical examples come from Wagner, while there are no Mozart or Bach examples in his book. His most intriguing speculation is that the perception and translation of musical movement at the scale he is interested in may be mediated by the vestibular organ, which controls body orientation and equilibrium. In support of this claim he cites scientific evidence from early physiological experiments. Furthermore, to illustrate the concrete instantiation of different movement types in music performance, Truslit presents recorded sound examples as well as some measurements of their acoustic microstructure. Although his empirical contribution is fairly negligible, his very modern theoretical ideas and the clarity and force with which they are presented must be greatly admired. (For an English translation of the gist of his book, see Repp, 1993.)

Two modern successors: Clynes and Todd

Despite the many interesting observations that these German pioneers, especially Becking and Truslit, have to offer to the modern reader, their work seems to have been largely forgotten. Some of their ideas may indeed be outmoded, but others are clearly relevant to more recent research on musical expression and performance. Among the small group of researchers active in this area, two seem particularly close in spirit to the German pioneers: Manfred Clynes and Neil Todd. Clynes was acquainted with Becking's work as he began in the 1970s to develop the concept of composers' "personal curves" further, making ingenious use of computer technology. Todd independently developed ideas resembling those of Truslit, without actually being aware of his work.

Over a number of years, Clynes (1977) developed the notion of *essentic forms*, dynamic shapes that characterize basic emotions. To measure these shapes, he devised a simple apparatus called the *sentograph*. It consists of a button sensitive to finger pressure in vertical and horizontal directions and a computer that registers the pressure over time and averages successive pressure cycles. Subjects who imagine certain basic emotions (love, anger, grief, etc.) while pressing rhythmically on the sentograph produce very different pressure curves for different emotions.

Clynes argues that meaning in music derives from essentic forms, which are conveyed by the musical structure (melody, rhythm) and microstructure (dynamics, agogics). The more closely an essentic form is approximated, the more beautiful and meaningful the music is perceived to be. This emotional "story", however, unfolds against the background of a fixed, repetitive, dynamic rhythmic pattern that represents the composer's individuality and "point of view". This is the composer's "inner pulse"—a concept clearly derived from Becking's theory of "personal curves". In his most recent writings, Clynes (1992) has referred to this as his "double stream theory" of musical expression.

The sentograph offered itself as a suitable instrument for measuring the essentic shapes of a piece of music as well as the composer's inner pulse. To assess the former, the (musically experienced) subject presses the button in synchrony with larger musical gestures or phrases while listening to or imagining a piece of music. To assess the latter, the subject presses more rapidly (about once per second) in synchrony with successive downbeats. These repeated pressure curves can then be averaged, yielding a stable average pulse shape. Such averaging is not easily possible with the longer essentic shapes, which may be one reason why Clynes has done little work to explore this aspect further.

To determine the shape of famous composers' inner pulses, Clynes used several outstanding musicians (including Pablo Casals, Rudolf Serkin, and himself) as subjects. They were asked to press rhythmically on the sentograph while imagining various works of Beethoven, Mozart, Schubert, and others. It was not a counterbalanced experiment—not every subject produced every composer's pulse, while some produced several pulse shapes for different pieces by the same composer. In any case, as can be seen in Figure 5, the average vertical pressure curves (see Clynes, 1977) show striking differences between composers (here, Beethoven, Mozart, and Schubert) and considerable agreement within composers across different subjects and different pieces.

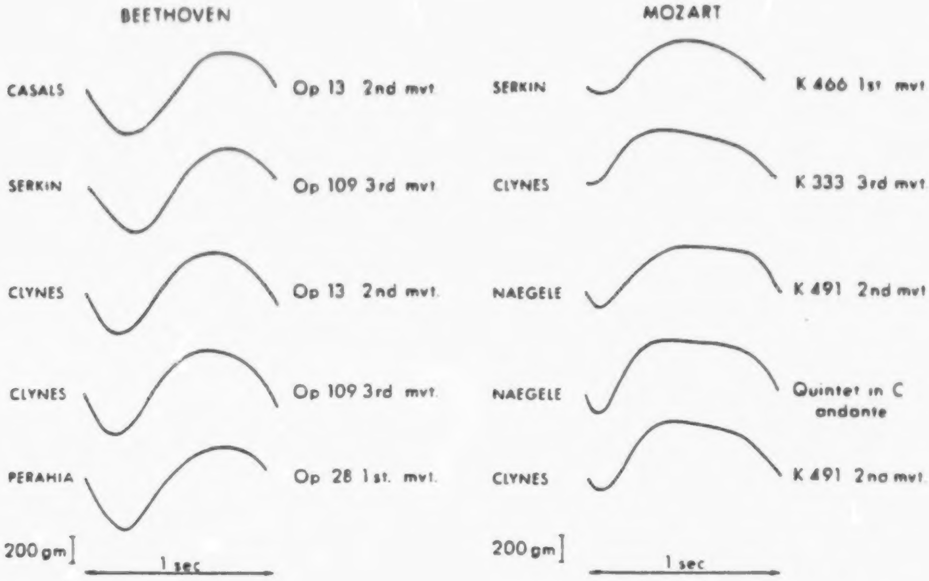


Figure 5. **Average sentograph responses (vertical pressure curves) of several distinguished musicians as they mentally rehearsed compositions by Beethoven and Mozart. (Reproduced from Clynes, 1969: 200, with permission of the publisher.)**

Clynes thus went one step beyond Becking by registering the "conducting" movements that Becking represented only schematically by means of graphs. Even though the finger movements on the sentograph are different from the baton-aided hand movements Becking had in mind, they seem to capture some of the composer-specific characteristics that Becking talked about. The main analogy between Becking's and Clynes's curves seems to lie in the onset time, relative speed, and depth of the downward movement.

Some years after his demonstration of composers' inner pulses on the sentograph, Clynes (1983) advanced towards an objectivization of the pulse concept. Although Becking had provided some hints towards the manner in which individual composers' pulses might be manifested on the musical surface of a performance, he basically thought of them as mental or "inner" phenomena. Clynes pursued the idea that composers' personal pulses must somehow be manifested in the expressive microstructure of an expert performance. Rather than analyzing the performances of great interpreters, he developed a computer program that enabled him to play back music with different agogic and dynamic patterns, repeated cyclically from bar to bar. Using himself as a listener and judge, he manipulated and refined these objective pulse patterns for various compositions of different composers, primarily Beethoven, Mozart, and Schubert. He eventually arrived at settings that he found optimally appropriate for each composer; these patterns were quite different across composers but seemed to fit different compositions by the same composer. They could be specified numerically in terms of the relative amplitudes and durations of the tones within a metric cycle. Subsequently, Clynes (1986) expanded his scheme to encompass one or two higher levels within which the basic pulse cycles are nested, and which in turn exhibit the temporal and dynamic relationships of the composer's inner pulse, so that the rhythmic surface pattern is a multiplicative combination of higher- and lower-level pulse parameters.

These pulse patterns, then, represent Clynes's subjective judgment, which identifies his enterprise as being partially in the intellectual tradition of Sievers and Becking. What distinguishes it from its historic precedents, however, is that the pulses are *quantified* and hence open to empirical tests. Several attempts have been made to test the effectiveness of Clynes's specifications in conveying the composer's individuality to unbiased listeners. The method was to generate computer performances of several composers' pieces with each composer's pulse, in a factorial design, and to see whether listeners prefer the performances with the "appropriate" pulse over the others.Several experiments by Thompson (1989) and by me (Repp, 1989, 1990a) have yielded mixed results, but the most recent study, conducted by Clynes himself (in press), provided unambiguous evidence that highly trained musicians prefer the appropriate composers' pulses over inappropriate pulses in computer-generated performances. However, questions remain about how a composer's inner pulse is manifested in human performances, where many factors besides the composer's individuality may affect the expressive microstructure (see Repp, 1990b; 1992).

In these studies, the emphasis was on the quantification and perceptual evaluation of cyclic pulse patterns, not so much on their relation to physical movement. Clynes and Walker (1982) addressed this latter point by investigating the biological "transfer function" between rhythmic sound patterns and the rhythmic movement of a human listener. The subject pressed on a sentograph while listening to cyclic repetitions of two tones having variable onset times, durations, and amplitudes. The resulting averaged pressure curves varied systematically with the sound patterns presented. For example, the downward movement of the finger, which usually accompanied the louder of the two tones, depended on the temporal separation of the softer tone from the louder tone. The timing of the upward movement depended on tone duration: Patterns of long tones resulted in smooth, round movements, whereas patterns of short tones (with long gaps in between) induced sharp, angular movements.

To relate these results, obtained with arbitrary rhythmic patterns, to the hypothesized pulse patterns of actual music, Clynes and Walker matched two-tone patterns to synchronously played music. They adjusted the physical parameters of the two tones until they perceived a congruence with the musical rhythm. Subsequently, they had subjects press the sentograph when listening either to the music or to the matched two-tone "sound pulse". Figure 6 shows that there was a significant similarity between these motoric responses, indicating that the simple two-tone pulse patterns captured the rhythmic pulse of the acoustically much more complex music.

Clynes's theories and research (of which I have provided here only a brief glimpse) represent a highly original and important contribution to music psychology. However, his observations are in need of extension and replication in other laboratories, as they are often based on very limited data. I find it regrettable that so few researchers have pursued the intriguing avenues opened up by this exceptionally creative mind in our midst.

While Clynes was inspired by the ideas of Becking, Todd is in some ways the intellectual heir of Truslit. The most obvious coincidence is both authors' hypothesis that the perception of musical motion may be mediated somehow by the vestibular system (Todd, 1992a, 1992b, 1992c; 1993). Although there is little evidence that vestibular stimulation actually occurs in ordinary music listening conditions, perhaps this is not really necessary: The sound patterns that characterize body movement could be recognized at an abstract auditory or cognitive level. They may be the very same as those that, under certain extreme conditions (e.g., in very loud music), can evoke vestibular sensations.

Like Truslit, Todd is concerned primarily with motion at the level of the whole body, rather than of the limbs or fingers. He, as did Truslit before him, appeals to physiological evidence for two distinct motor systems, the ventromedial and lateral systems (Todd, 1992b). The former controls body posture and motion, and is closely linked with the vestibular system. Since larger masses are to be moved, the movements are slower than those possible with feet, hands, and fingers, which are controlled by the lateral system. Typically their cycles extend over several seconds, whereas the pulse microstructure studied by Clynes (and executed by finger pressure on the sentograph) is contained within cycles of roughly 1 s duration that may be nested within the larger cycles described by Truslit and Todd. Recently, Todd (1992c; Todd, Clarke, & Davidson, 1993) has begun to study the motoric instantiation of these larger cycles in the "expressive body sway" of performers. His preliminary data indicate that pianists' head movements are synchronized with expressive tempo fluctuations in the music, such that tempo minima coincide with turning points in the head movement. Observations such as these have led Todd to propose that expressive variation in tempo and in the correlated dynamics may be a representation of self-motion. Clearly, such a representation has the potential of inducing actual or imaginary motion of a similar kind in a listener/observer.

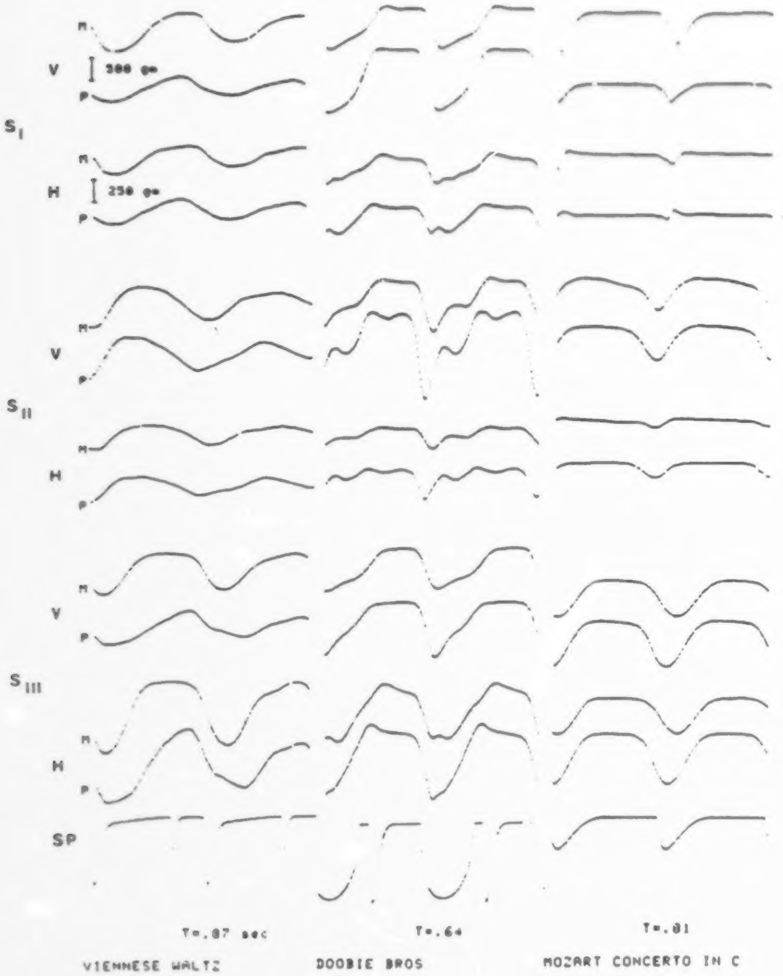


Figure 6. **Vertical (V) and horizontal (H) sentograph pressure curves for three subjects (S_I, S_{II}, S_{III}) who listened either to music (M) or to a simple sound pulse (P) matched to the music. Each column represents a different piece of music. The amplitude envelope of the sound pulse (SP) is shown at the bottom. (Reproduced from Clynes & Walker, 1982: 211, with permission of the publisher.)**

Concerning the tempo variations in performances, Todd (1992a, 1992d) has presented evidence that they are a linear function of real time. In other words, expressive timing consists of alternating phases of constant acceleration and deceleration, one cycle typically corresponding to a musical group or phrase. Listeners also seem to prefer performances whose timing follows this rule, although more extensive perceptual tests remain to be done. Constant acceleration or deceleration seems to characterize various forms of physical and biological motion, so that music having this property would seem an optimal stimulus for the perception and induction of motion. Todd (1992a) has also begun to investigate the way in which changes in musical dynamics go along with changes in timing and has devised a system for the automatic extraction of hierarchical rhythmic structure from the amplitude envelope of the acoustic signal (Todd, 1994). This is exciting work at the cutting edge of contemporary research on music performance.

Research on Biological Motion

Evidence for constraints on natural motion comes from research on human motor control. There is one body of research that seems particularly relevant to me, which is due to Paolo Viviani and his collaborators (see Viviani, 1990; Viviani & Laissard, 1991). Over the last decade, they have investigated the constraints that link the geometry and the kinematics of guided hand movements. The movements in question involved the drawing or tracking of ellipses or of more complex curvilinear paths. The consistent finding has been that, within a single coherent movement, velocity varies as a power function of trajectory curvature (Viviani & Terzuolo, 1982; Viviani & Cenzato, 1985; Viviani & Schneider, 1991). In other words, the greater the local curvature, the slower the movement. Viviani, Campadelli, and Mounod (1987) have demonstrated that subjects are unable to track accurately a light point moving at a constant velocity around an elliptic path, whereas the task is easy when the target velocity changes with curvature according to the power function. It has also been shown that dynamic visual stimuli of the latter kind are judged by observers to represent constant velocity, whereas elliptic stimuli exhibiting constant velocity seem to vary in velocity (Viviani & Stucchi, 1992).

The *spatio-temporal coupling* described in this research on biological movement enhances considerably the scientific respectability of the technique of "accompanying movements" developed by Sievers, Becking, and particularly by Truslit. If spatial trajectory determines the velocity profile, then a particular velocity profile also implies a spatial trajectory of a particular kind. What Truslit evidently did was to convert the velocity information available in the temporal and dynamic microstructure of music into arm movements with a matching spatial trajectory whose direction also took the melodic pitch contour into account. In his book he describes how, with practice, a close subjective match between the spatial trajectory and the auditory information can be achieved. What seemed like a highly idiosyncratic method at first may in fact have a solid foundation in the constraints of biological motion.

Conclusions

I conclude from this very limited survey that music, by virtue of its temporal and dynamic microstructure, has the potential to represent forms of natural motion and to elicit corresponding movements in a human listener. While a rigid rhythm may elicit only foot tapping or finger snapping, an expressively modulated structure can specify movements with complex spatial trajectories that, for the purpose of demonstration and analysis, can be realized as guided movements of the limbs or the whole body. However, execution of such movements is not necessary to appreciate the motion information: Experienced listeners, at least, can judge by ear whether the musical motion is natural or awkward, and they can move along with the music inwardly, as it were. An aesthetically satisfying performance presumably is one whose microstructure satisfies basic constraints of biological motion while also being responsive to the structural and stylistic requirements of the composition.

REFERENCES

Becking, G. (1928). *Der musikalische Rhythmus als Erkenntnisquelle.* Augsburg: Benno Filser.

Clynes, M. (1969). Toward a theory of man: Precision of essentic form in living communication. In K. Leibovic & J. Eccles (Eds.), *Information processing in the nervous system* (pp. 177-206). New York: Springer-Verlag.

Clynes, M. (1977). Sentics: The touch of the emotions. New York: Doubleday.

Clynes, M. (1983). Expressive microstructure in music, linked to living qualities. In J. Sundberg (Ed.), *Studies of music performance* (pp. 76-181). Stockholm: Royal Swedish Academy of Music (Publication No. 39).

Clynes, M. (1986). Generative principles of musical thought: Integration of micro-structure with structure. *Communication and Cognition (CC-AI)*, 3, 185-223.

Clynes, M. (1992). Time-forms, Nature's generators and communicators of emotion. In *Proceedings of the IEEE International Workshop on Robot and Human Communication, Tokyo* (pp. 1-14).

Clynes, M. (in press). Composers' pulses are liked most by the best musicians. *Cognition.*

Clynes, M., & Walker, J. (1982). Neurobiologic functions of rhythm, time, and pulse in music. In M. Clynes (Ed.), *Music, mind, and brain* (pp. 171-216). New York: Plenum Press.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

Langer, S. (1953). *Feeling and form.* New York: Scribner.

Nohl, H. (1920). *Stil und Weltanschauung.* Jena: Eugen Diederichs.

Repp, B. H. (1989). Expressive microstructure in music: A preliminary perceptual assessment of four composers' "pulses". *Music Perception*, 6, 243-274.

Repp, B. H. (1990a). Further perceptual evaluations of pulse microstructure in computer performances of classical piano music. *Music Perception*, 8, 1-33.

Repp, B. H. (1990b). Patterns of expressive timing in performances of a Beethoven minuet by nineteen famous pianists. *Journal of the Acoustical Society of America*, 88, 622-641.

Repp, B. H. (1992). Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's "Träumerei". *Journal of the Acoustical Society of America*, 92, 2546-2568.

Repp, B. H. (1993). Music as motion: A synopsis of Alexander Truslit's (1938) "Gestaltung und Bewegung in der Musik". *Psychology of Music*, 21, 48-72.

Rutz, O. (1911). *Musik, Wort und Körper als Gemütsausdruck*. Leipzig.

Rutz, O. (1922). *Sprache, Gesang und Körperhaltung* (2nd edition). München.

Sessions, R. (1950). *The musical experience of composer, performer, listener*. Princeton: Princeton University Press.

Sievers, E. (1924). *Ziele und Wege der Schallanalyse*. Heidelberg: Carl Winter's Universitätsbuchhandlung.

Thompson, W. F. (1989). Composer-specific aspects of musical performance: An evaluation of Clynes's theory of pulse for performances of Mozart and Beethoven. *Music Perception*, 7, 15-42.

Todd, N. P. McA. (1992a). The dynamics of dynamics: A model of musical expression. *Journal of the Acoustical Society of America*, 91, 3540-3550.

Todd, N. P. McA. (1992b). Music and motion: A personal view. In C. Auxiette, C. Drake, and C. Gérard (Eds.), *Proceedings of the Fourth Rhythm Workshop: Rhythm Perception and Production* (pp. 123-128). Bourges, France.

Todd, N. P. McA. (1992c). The communication of self-motion in musical expression. *Proceedings of the International Workshop on Man-Machine Studies in Live Performance* (pp. 151-162). Pisa, Italy: CNUCE/CNR.

Todd, N. P. McA. (1992d). The kinematics of musical expression. Submitted for publication.

Todd, N. P. McA. (1993). Vestibular feedback in musical performance: Response to *Somatosensory Feedback in Musical Performance* (edited by Sundberg and Verrillo). *Music Perception*, 10, 379-382.

Todd, N. P. McA. (1994). The auditory "primal sketch": A multiscale model of rhythmic grouping. *Journal of New Music Research*, 23 (forthcoming).

Todd, N. P. McA., Clarke, E. F., & Davidson, J. (1993). The representation of self-motion in expressive performance. Paper presented at the Annual Conference of the Society for Music Perception and Cognition, June 16-19, University of Pennsylvania, Philadelphia.

Truslit, A. (1938). *Gestaltung und Bewegung in der Musik*. Berlin-Lichterfelde: Chr. Friedrich Vieweg.

Ungeheuer, G. (1964). Die Schallanalyse von Sievers. *Zeitschrift für Mundartenforschung*, 31, 97-124.

Viviani, P. (1990). Common factors in the control of free and constrained movements. In M. Jeannerod (Ed.), *Attention and Performance XIII* (pp. 345-373). Hillsdale, NJ: Erlbaum.

Viviani, P., Campadelli, P., and Mounoud, P. (1987). Visuo-manual pursuit tracking of human two-dimensional movements. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 62-78.

Viviani, P., and Cenzato, M. (1985). Segmentation and coupling in complex movements. *Journal of Experimental Psychology: Human Perception and Performance*, 11, 828-845.

Viviani, P., and Laissard, G. (1991). Timing control in motor sequences. In J. Fagard & P. H. Wolff (Eds.), *The Development of Timing Control and Temporal Organization in Coordinated Action* (pp. 1-36). Amsterdam: Elsevier.

Viviani, P., and Schneider, R. (1991). A developmental study of the relationship between geometry and kinematics in drawing movements. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 198-218.

Viviani, P., and Stucchi, N. (1992). Biological movements look uniform: Evidence of motor-perceptual interactions. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 603-623.

Viviani, P., and Terzuolo, C. (1982). Trajectory determines movement dynamics. *Neuroscience*, 7, 431-437.

Zuckerkandl, V. (1973). *Sound and symbol: Music and the external world* (W. R. Trask, trans.). Princeton: Princeton University Press. (Original work published 1956.)

FOOTNOTE

*Invited paper presented at the Stockholm Music Acoustics Conference (SMAC93) in August 1993. This version minus the figures and related text will appear in the SMAC93 Proceedings. A revised version including the figures constitutes the second half of a chapter by Patrick Shove and Bruno H. Repp, "Musical Motion and Performance: Theoretical and Empirical Perspectives", to appear in *Performance Studies*, edited by John Rink (Cambridge, UK: Cambridge University Press).

Haskins Laboratories Status Report on Speech Research
1993, SR-114, 179-184

A Review of P. Downing, S. D. Lima, & M. Noonan (Eds.), *The Linguistics of Literacy*[*]

Ignatius G. Mattingly[†]

Linguists have always been suspicious of traditional orthographies. After all, a traditional orthography directly competes with the linguist, offering its own morphophonological analysis of the language in question. This state of affairs is the more painful because it often happens that an orthography is, by any reasonable linguistic criteria, totally unsuited to the language it transcribes, and yet its users seem perfectly happy with it, and resist all attempts to simplify or rationalize it. The linguist is in the position of a highly-trained physician unable to persuade patients to give up their ineffective and unscientific folk nostrums.

It is thus no surprise that a book growing out of a Symposium whose theme was "the relationship between linguistics and literacy" (p. ix) should provide further evidence that this relationship is an uneasy one. The book includes fifteen papers presented at the Seventeenth Annual University of Wisconsin-Milwaukee Linguistics Symposium in Milwaukee, April 8-10, 1988. The editors have grouped the papers into three parts: "Written Language and Spoken Language Compared," "Orthographic Systems," and "Psychology of Orthography." A fourth part, "Consequences of Literacy," consists of a sixteenth paper, not presented at the Symposium, by Walter J. Ong.

Part I, comparing speech and writing, includes papers by Cecilia E. Ford, Wallace Chafe, Deborah Tannen, and Eleanor Berry. Ford compares the intonation of adverbial clauses in samples of conversation with the punctuation of such clauses in samples of Freshman writing. She finds that in the conversation samples, temporal clauses are more likely than conditional or causal clauses to be part of the same breath group as the main clause, or to be preceded by intonation contours signaling incompletion, rather than being preceded by contours signaling completion. In the written samples, similarly, temporal clauses are more likely than conditional or causal clauses to be connected to the main clause with no punctuation, rather than being separated by commas, periods, or dashes. Thus, in this respect at least, writing parallels speech.

Tannen considers another similarity between writing and speech: Both literary artists and conversationalists make their effects by introducing striking details that may be logically superfluous to the apparent message. In support of this claim, she provides many impressive examples from both domains. Of course, the point is hardly novel. Schoolteachers and literary critics have always stressed the importance of imagery in literature, and literary history records the struggles of successive generations of poets to restore to poetic language the concreteness of common speech.

The attempts of American Modernist poets (Frost, Eliot, and Williams) in this direction are discussed by Berry. She finds that despite the professed allegiance of these poets to conversational speech, art keeps creeping into their work, even in poems that purport to be conversational narratives. Thus, repetitions, hesitations, replacements of words, and afterthoughts are much less frequent in their poetry than in actual conversation, and when present are apt to have an obvious artistic function. Modernist poetry is not a literal transcription of common speech, but a highly organized and densely structured rearrangement of it. Thus, for Berry, it is a *difference* between written and spoken language that is of interest.

Chafe offers evidence of another sort of difference, pointing to certain constraints on spoken language absent in written language. In his conversational samples, speakers present no more than one "new idea" per breath group, and this new idea is never embodied in the grammatical subject. But parallel constraints do not, of course, apply to written clauses, nor are these constraints adhered to in oral reading, even when a clause is divided into two or more breath groups. This freedom of written language he attributes to "(1) the reduced burden on readers made possible by their role as consumers rather than producers and (2) the freedom of readers to set their own pace" (p. 27).

All four of these papers are open to the objection that comparisons between speaking and writing are confounded with comparisons between modes of discourse (dialog, narration, exposition, persuasion, etc.). Instead of comparing speech and writing within a particular mode, they compare spoken dialog with writing in some other mode. A justification for this way of proceeding is offered by Chafe: "...ordinary casual conversation has a special position as the prototypical use of spoken language.... It is helpful to be able to identify a use of language on which we can anchor our study, and with relation to which we can interpret other, less prototypical uses" (p. 19). But this is not entirely convincing, particularly as it turns out that "written language seems not to offer anything comparable [to conversation] as a prototype" (p. 23). But by assuming that only conversational speech is truly prototypical, Chafe is able to scoff at linguists who discuss examples not apt to be found there, like Sapir's (1921) *The farmer kills the duckling* (p. 82) or an unnamed linguist's *The managing of an office by Peter is liked by John*. Yet surely the interesting point is that, despite their alleged unutterability, these sentences are perfectly comprehensible and grammatically acceptable.

Part II, on writing systems, includes papers by Mark Aronoff, Peter T. Daniels, Alice Faber, Janine Scancarelli, Ronald P. Schaefer, and James D. McCawley. The first three authors consider the relation between phonological and orthographic units. Aronoff calls our attention to Baron Massias, an obscure nineteenth-century French philosopher who held that "writing, specifically alphabetic writing, lies at the heart of language" (p. 73). Although linguists would now agree that writing is just a secondary system, some of them, according to Aronoff (and so also Faber and, in Part III, Bruce Derwing), have inadvertently fallen into a way of thinking akin to that of Massias. The phonemic segment is a misconception to which they have been unwittingly led by their experience with alphabetic writing. This is not, as he acknowledges, a new idea, but "good ideas sometimes bear repeating" (fn. 2, p. 81). While phonemic segments are fair game, Aronoff is hardly justified in deriding Saussure, Sapir, and Chomsky and Halle as having been "caught in the web of their own orthography" (p. 81). His evidence for this in Saussure's case is a passage from Baskin's translation of the *Course in General Linguistics* (1959, pp. 38-39) in which Saussure says that in Greek writing, letters correspond to auditory beats. But the passage shows only that Saussure believed that the existence of alphabetic writing was consistent with his notions about phonemic segments, not that his personal experience with alphabetic writing had shaped these notions. The evidence from Sapir (1933) and from Chomsky and Halle (1968) is unconvincing not only for similar reasons, but also because, in the passages cited from these authors, what is at issue is not the segmental character of phonemes but their level of abstraction.

It is doubtless true that these linguists, like most literate Westerners, originally acquired their notion of the phonemic segment through exposure to an alphabetic orthography. But they did not accept this notion uncritically or unreflectingly; they considered a great many linguistic data and, rightly or wrongly, determined that a segmental analysis best accommodated the observed regularities. Moreover, Chomsky and Halle, at any rate, were surely well aware of various counterproposals, such as those of Harris (1944) and Firth (1967), even though they did not yet see how to reconcile these proposals with the evidence for a segmental account. Aronoff views the recent trend toward nonlinear models in phonological theory as belatedly liberating phonology from the grip of the alphabet. But these new phonological models did not arise in a nonliterate culture, or even in one using a syllabary, but rather in the same alphabet-ridden culture that had produced segmental phonology; some of them, indeed, were encouraged by Halle himself, and they are more reasonably viewed as generalizations of the Chomsky and Halle (1968) model than as rejections of it.

For Daniels, the syllable is "the most salient unit of speech" (p. 84) and the Sumerian, Chinese, and Mayan syllabic writing systems could be invented because morphemes in these languages were generally monosyllabic. On the other hand, alphabets, being based on the phoneme, are quite unnatural. However, Daniels' rambling paper does not confine itself to these matters. He finds room for a great deal of material of questionable relevance, for the introduction of novel terminology of which he then makes no use, and for the accusation that when Martin Joos reprinted W. F. Twaddell's *On Defining the Phoneme* (1935) in *Readings in Linguistics I* (1966), he omitted passages on acoustic phonetics showing that phonemes were not manifest in the speech signal, thereby undermining Sapir's (1933) view that the phoneme was a mentalistic abstraction. This slur on Joos's scholarly integrity is unjust and reckless. A comparison of Twaddell (1935) with Joos's abridged version (Twaddell, 1966) shows that the omitted passages (1935, pp. 35-36, cf. 1966, p.68; 1935, pp. 56-57, cf. 1966, p. 77) are quite adequately summarized in Twaddell (1966), either by Joos or by Twaddell himself. Anyway, Joos himself, a pioneer in the spectrographic analysis of speech, certainly did not share the simplistic view of the acoustic status of phonemic segments held by older American structuralists like Bloomfield (1933), as his classic monograph, *Acoustic Phonetics* (1948) testifies.

Faber sets herself the task of demonstrating how, given that phonology itself is not segmental, segmental writing could have arisen. She argues that it is not justifiable to attribute segmental awareness to the Phoenicians; they must have been aware of the consonants that they actually transcribed, but not necessarily of the different vocalic patterns, interdigitated with the consonants in speech, that they did not transcribe. Thus, there is no reason to believe they would have analyzed a syllable such as /ba/ into two successive segments. To account for the emergence of the *plene* Greek alphabet, she adopts Sampson's (1985) proposal that the Greeks heard /ʔalpa/, /ħe/, /yoda/, and /ʕɑyna/, the Canaanite letter-names for the consonants /ʔ/, /ħ/, /y/, and /ʕ/, as alpa/, /e/, /ioda/, and /ɑyna/, because those consonants do not occur in Greek. Therefore, they took the corresponding letters to stand instead for the vowels /a/, /e/, /i/, /ɑ/, and could interpret the Canaanite system as fully alphabetic. Thus segmental awareness arose in the Greeks for the same reason it has in all their successors: as a result of exposure to what appeared to be a segmental writing system. There is no need to assume on anyone's part a prior, phonologically rather than orthographically based, segmental awareness.

One cannot but admire Faber's ingenuity in avoiding an appeal to awareness of phonological segments, but certain questions arise. How would she explain the later Semitic writing systems for Aramaic and Hebrew, in which yod, waw and aleph were sometimes used to represent vowels (Cross & Freedman, 1952)? Hadn't segmental awareness crept in somehow by this stage? Or again, on Sampson's account, the Greeks would have seen Canaanite writing as a system in which only a minority of the vowels, those that were apparently syllable-initial, were transcribed. Didn't it require some prior segmental awareness on their part to generalize the principle to vowels in other positions?

Scancarelli takes a close look at Sequoyah's Cherokee syllabary. This writing system is not as ideal as it is often said to be. For example, separate signs are in a few cases provided for two CV syllables contrasting in aspiration, but not in many others. To account for this, she suggests, very plausibly, that Sequoyah minimized his inventory of symbols by assigning separate signs to the members of such a pair only when it seemed to him that their contrast carried a high functional load.

Schaefer describes the various strategies employed by naive users of the orthography devised for Emai (an Edoid language of Southern Nigeria) to transcribe phrases containing elisions of word-final vowels. These writers never represent the quality of the elided vowel, but sometimes they preserve the lexical shape of the word that contained it, of the word following, or of both. Thus *vbae eo* > *vba eo* and *vbi ogo* > *vbogo*, but *vbi ean* > *vbe an*, *eli eami* > *ele ‘ami*, *re obo* > *ro obo*, *ze obo* > *zi obo*. Schaefer attributes these patterns to greater awareness of lexical shape than of phonemic shape. This may be so, but the fact that the writers did consistently indicate the elision suggests a degree of phonological awareness. And what more obvious indication is there than the omission of the letter for the elided vowel?

The last essay in this section, by McCawley, is a discussion of musical and mathematical notation. It is extremely lucid and at times brilliant, but seems out of place in this book concerned with linguistic writing and natural language. McCawley makes intriguing comparisons between the structures of these notations and linguistic structure, showing, for example, how music indicates constituents with beams and ligatures. He is rather less convincing when he suggests that the correspondence between the position of the notes on a staff and pitch height is a metaphor; why is this not just iconicity? Nor does there seem much point in regarding sin^2x as an "optional transformation" of $(sin\ x)^2$. It would have been of greater value and relevance to compare mathematical or musical notation with linguistic writing, a notation for language, rather than with language itself. It is of considerable interest, for example, that, unlike these other notations, conventional linguistic writing does *not* indicate constituent structure.

Part III, on the psychology of orthography, includes papers by Bruce Derwing, John Ohala, Laurie Feldman, Ram Frost, and J. Ronayne Cowan. On the basis of subjects' performance in such tasks as phonetic similarity judgment and segment counting, Derwing argues convincingly that the phonology of literate speakers is heavily impacted by their orthographic experience and that writing and reading cannot be set aside as merely parasitic on speaking and listening. (Is there then something to be said for Massias' views after all?) "This evidence suggests a kind of lexicon in which the phonological and orthographic representations are not sharply separated" (p. 197). Derwing also wants to argue, like Aronoff and Faber, that the orthography has beguiled linguists into setting up a unit, the phoneme, that is psychologically unreal. But these two views seem almost contradictory. Derwing seems to be saying, on the one hand, that the orthography has far more profound psycholinguistic effects than is commonly supposed, and on the other, that the units it implies are psycholinguistically irrelevant!

Ohala proposes a "cost-benefit" evaluation of generative phonology and argues that for most speakers, the cognitive cost (the effort required for phonological analysis) outweighs the benefit (identification of morphemes recurring in different form in different words). "Different pronunciations of the same morpheme...are largely nonfunctional and are rather to be viewed as an unfortunate but inevitable consequence of the ravages of sound change" (p. 229). He offers data—spelling errors and naive judgments whether pairs of words are historically related—showing that speakers do not, in fact, carry out phonological analysis consistently. According to Ohala, they need not and for the most part do not set up underlying forms. All that they really require are a few "cut-and-paste rules," i. e., analogies. Generative phonology is just disguised diachronic phonology.

Confronted with this hardnosed stance, the generative phonologist might respond that he is concerned with the phonology of *ideal* speaker-hearers, for whom the only relevant "cost" is the complexity of the phonology. He would willingly concede that, no doubt for the various reasons Ohala gives, this ideal is realized very imperfectly in actual speakers. For someone who insists on doing traditional armchair phonology, the only possible alternative to this position is that of Twaddell (1935): The phoneme is a fiction.

But perhaps the prospect from the armchair is not so bleak after all. Feldman reviews the results of a number of repetition priming experiments carried out by her and her colleagues. In this technique, two related items are presented separately for lexical decision, with other trials intervening. For some types of relation, the second item is responded to faster than when no related item has preceded. Fowler, Napps, and Feldman (1985) found such a facilitating effect for priming with morphologically related words, and the effect was just as great when the pronunciation of the morpheme differed in the two words (*heal, health*), or both its spelling and its pronunciation differed (*clear, clarify*), as when spelling and pronunciation did not change (*heal, healer; clear, clearly*). This finding surely argues for the psychological reality of the constructs of generative phonology, at least for literate speakers. Ohala does refer to Fowler et al. (1985) and to Feldman's paper, but only to remark that "repetition priming...appears capable of providing behavioral evidence relevant to the issue" (p. 226).

Frost summarizes evidence for the effects of "orthographic depth" (Klima, 1972) on the reading process. An orthography is deep, according to Klima, if it appeals to the more abstract level of morphophonology where morphemes have a constant shape, rather than to a level nearer the surface, such as the phonemic level of structuralist phonology. English orthography is thus deep, that of Hebrew even deeper, but that of Serbo-Croatian is shallow. Certain experimental tasks, for example, naming, are performed faster and more accurately for shallow than for deep orthographies (Frost, Katz, & Bentin, 1987), and it is clear that the dimension of orthographic depth has some psychological reality. Frost is careful, however, not to claim more for orthographic depth than is warranted. It is not to be concluded that deep orthographies are processed in some radically different, possibly more "visual" way than shallow ones.

In perhaps the only paper in this collection that has something positive to say about orthographies, Cowan offers evidence that American students make good use of the orthography of the second language they are learning. This is reflected both in the errors they make and in their greater ability to retain vocabulary words if the orthographic form of the word is presented along with spoken form.

Moreover (though Cowan does not employ this terminology), shallow orthographies are more helpful than deep ones. But is this state of affairs really desirable? Apparently, these language learners, rather than confronting a new and strange phonology, are attempting to assimilate it as far as possible to their native phonology, and the orthography helps them to do this. Should they be allowed this crutch, if they are really to learn a foreign language?

In Part IV, Ong charges that writing has cut us off from the world of "primary oral culture." Writing separates the known from the knower, interpretation from data, word from sound, source from recipient, language from the plenum of existence, past from present, and so on. Before the advent of writing, each of these oppositions was a unity. Literacy has, indeed, some compensations: We can be objective and consciously aware of things in a way that was not possible before writing. Ong even grants, as Aronoff would not, that "writing can distance us from writing itself... Writing has the power to liberate us more and more from the chirographic bias and confusion it creates, though complete liberation is impossible" (p. 316).

But surely Ong unduly idealizes and oversimplifies oral cultures. Can we really be sure that they all are "basically conservative" (p. 295), "incapable of linear analysis"(p. 298), "mobile, warm, personally interactive" (p. 299), and that they all "view everything in terms of interpersonal struggle" (p. 298) and "use words less for information and more for optional, interpersonal purposes" (p. 306)? Ong's oral culture is unreal, a lost Eden to be nostalgically recalled: "Of course, the original innocence of the pristine empathetic identification can never be repossessed directly" (p. 317).

Orthography, particularly alphabetic orthography, it seems, has much to answer for. It is less natural than conversational speech (Chafe, Berry), it misleads linguists (Aronoff, Faber, Derwing), it relies on an unnatural unit (Daniels), it corrupts one's phonology (Derwing), and it has cut us off forever from primary oral culture (Ong). But it is hard to imagine giving it up. We are all hooked at an early age, and while our heads tell us that orthographies are merely secondary systems, our hearts say that Baron Massias was right.

REFERENCES

Bloomfield, L. (1933). *Language*. New York. Henry Holt.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Cross, F. M., & Freedman, D. N. (1952). *Early Hebrew orthography/A study of the epigraphic evidence*. New Haven, CT: American Oriental Society.

Firth, J. R. (1967). *Papers in linguistics*. Oxford: Oxford University Press.

Fowler, C. A., Napps, S. E., & Feldman, L. B. (1987). Relations among regular and irregular morphologically related words in the lexicon as revealed by repetition priming. *Memory & Cognition, 13*, 241-255.

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 104-114.

Harris, Z. (1944). Simultaneous components in phonology. *Language, 20*, 181-205.

Joos, M. (1948). *Acoustic phonetics*. (Language Monographs, No. 23). Baltimore: Waverly Press, for the Linguistic Society of America.

Joos, M. (1966). *Readings in linguistics I*. 4th ed. Chicago: University of Chicago Press. [First published 1957].

Klima, E. S. (1972). How alphabets might reflect language. In J. F. Kavanagh & I. G. Mattingly (Eds.) *Language by ear and by eye*. (pp. 57-80). Cambridge, MA: MIT Press.

Sampson, G. (1985). *Writing systems.*. Stanford, CA: Stanford University Press.

Sapir, E. (1921). *Language*. New York: Harcourt Brace.

Sapir, E. (1933). La réalité psychologique des phonèmes. *Journal de Psychologie Normal et Pathologique, 30*, 247-65. Translated in D. Mandelbaum (Ed.), *Selected writings of Edward Sapir in language, culture, and personality* (pp. 46-60), Berkeley: University of California Press, 1963.

Saussure, F. (1959). *Course in general linguistics*. W. Baskin (Tr.) New York: Philosophical Library.

Twaddell, W. F. (1935). *On defining the phoneme*. (Language Monographs, No. 16). Baltimore: Waverley Press, for the Linguistic Society of America.

Twaddell, W. F. (1966). On defining the phoneme. In M. Joos(Ed.), *Readings in linguistics I*. 4th ed (pp. 55-80). Chicago: University of Chicago Press. [First published 1957].

FOOTNOTES

[*] Amsterdam/Philadelphia: John Benjamins Publishing Company, 1992, xx, 334 pp. This review appears in *Language and Speech, 37*, 87-93 (1994).

[†] Also University of Connecticut.

Appendix

SR #	Report Date	NTIS #	ERIC #
SR-21/22	January-June 1970	AD 719382	ED 044-679
SR-23	July-September 1970	AD 723586	ED 052-654
SR-24	October-December 1970	AD 727616	ED 052-653
SR-25/26	January-June 1971	AD 730013	ED 056-560
SR-27	July-September 1971	AD 749339	ED 071-533
SR-28	October-December 1971	AD 742140	ED 061-837
SR-29/30	January-June 1972	AD 750001	ED 071-484
SR-31/32	July-December 1972	AD 757954	ED 077-285
SR-33	January-March 1973	AD 762373	ED 081-263
SR-34	April-June 1973	AD 766178	ED 081-295
SR-35/36	July-December 1973	AD 774799	ED 094-444
SR-37/38	January-June 1974	AD 783548	ED 094-445
SR-39/40	July-December 1974	AD A007342	ED 102-633
SR-41	January-March 1975	AD A013325	ED 109-722
SR-42/43	April-September 1975	AD A018369	ED 117-770
SR-44	October-December 1975	AD A023059	ED 119-273
SR-45/46	January-June 1976	AD A026196	ED 123-678
SR-47	July-September 1976	AD A031789	ED 128-870
SR-48	October-December 1976	AD A036735	ED 135-028
SR-49	January-March 1977	AD A041460	ED 141-864
SR-50	April-June 1977	AD A044820	ED 144-138
SR-51/52	July-December 1977	AD A049215	ED 147-892
SR-53	January-March 1978	AD A055853	ED 155-760
SR-54	April-June 1978	AD A067070	ED 161-096
SR-55/56	July-December 1978	AD A065575	ED 166-757
SR-57	January-March 1979	AD A083179	ED 170-823
SR-58	April-June 1979	AD A077663	ED 178-967
SR-59/60	July-December 1979	AD A082034	ED 181-525
SR-61	January-March 1980	AD A085320	ED 185-636
SR-62	April-June 1980	AD A095062	ED 196-099
SR-63/64	July-December 1980	AD A095860	ED 197-416
SR-65	January-March 1981	AD A099958	ED 201-022
SR-66	April-June 1981	AD A105090	ED 206-038
SR-67/68	July-December 1981	AD A111385	ED 212-010
SR-69	January-March 1982	AD A120819	ED 214-226
SR-70	April-June 1982	AD A119426	ED 219-834
SR-71/72	July-December 1982	AD A124596	ED 225-212
SR-73	January-March 1983	AD A129713	ED 229-816
SR-74/75	April-September 1983	AD A136416	ED 236-753
SR-76	October-December 1983	AD A140176	ED 241-973
SR-77/78	January-June 1984	AD A145585	ED 247-626
SR-79/80	July-December 1984	AD A151035	ED 252-907
SR-81	January-March 1985	AD A156294	ED 257-159
SR-82/83	April-September 1985	AD A165084	ED 266-508
SR-84	October-December 1985	AD A168819	ED 270-831
SR-85	January-March 1986	AD A173677	ED 274-022
SR-86/87	April-September 1986	AD A176816	ED 278-066
SR-88	October-December 1986	PB 88-244256	ED 282-278

SR-89/90	January-June 1987	PB 88-244314	ED 285-228
SR-91	July-September 1987	AD A192081	**
SR-92	October-December 1987	PB 88-246798	**
SR-93/94	January-June 1988	PB 89-108765	**
SR-95/96	July-December 1988	PB 89-155329	**
SR-97/98	January-June 1989	PB 90-121161	ED32-1317
SR-99/100	July-December 1989	PB 90-226143	ED32-1318
SR-101/102	January-June 1990	PB 91-138479	ED325-897
SR-103/104	July-December 1990	PB 91-172924	ED331-100
SR-105/106	January-June 1991	PB 92-105204	ED340-053
SR-107/108	July-December 1991	PB 92-160522	ED344-259
SR-109/110	January-June 1992	PB 93-142099	ED352594
SR-111/112	July-December 1992	PB 93-216018	ED359575
SR-113	January-March 1993		
SR-114	April-June1993		

AD numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22151

ED numbers may be ordered from:

ERIC Document Reproduction Service
Computer Microfilm Corporation (CMC)
3900 Wheeler Avenue
Alexandria, VA 22304-5110

In addition, Haskins Laboratories Status Report on Speech Research is abstracted in *Language and Language Behavior Abstracts*, P.O. Box 22206, San Diego, CA 92122

**Accession number not yet assigned

Haskins Laboratories Status Report on

Speech Research

SR-114
APRIL-JUNE 1993

Contents

- Evidence of Flexible Coding in Visual Word Recognition
  Kenneth R. Pugh, Karl Rexer, and Leonard Katz 1
- Horizontal and Vertical Views of Chinese Psycholinguistics
  Ignatius G. Mattingly 27
- Learning to Perceive the Sound Pattern of English
  Catherine T. Best 31
- The Perceptual Infrastructure of Early Phonological Development
  Alice Faber and Catherine Best 81
- The Stability of Distinctive Vowel Length in Thai
  Arthur S. Abramson 91
- Calibration, Validation, and Hardware-Software Modifications to the Carstens EMMA System
  Peter J. Alfonso, J. Robert Neely, P. H. H. M. van Lieshout, W. Hulstijn, and H .F. M. Peters 101
- Morphological Analysis and the Acquisition of Morphology and Syntax in Specifically-Language-Impaired Children
  Karen M. Smith-Lock 113
- Hemispheric Asymmetries in Adults' Perception of Infant Emotional Expressions
  Catherine T. Best, Jane S. Womer, and Heidi Fréya Queen 139
- On Determining the Basic Tempo of an Expressive Music Performance
  Bruno H. Repp 159
- Musical Motion: Some Historical and Contemporary Perspectives
  Bruno H. Repp 167
- A Review of P. Downing, S. D. Lima, & M. Noonan (Eds.), *The Linguistics of Literacy*
  Ignatius G. Mattingly 179

Appendix 185